Safety Regulation Group

Civil Aviation Authority

CAA PAPER 2002/05

Methods used to Evaluate the Effectiveness of Flightcrew CRM Training in the UK Aviation Industry

www.caa.co.uk

Safety Regulation Group

Civil Aviation Authority

CAA PAPER 2002/05

Methods used to Evaluate the Effectiveness of Flightcrew CRM Training in the UK Aviation Industry

Report prepared by Paul O'Connor, Rhona Flin, Georgina Fletcher, (Industrial Psychology Group, Aberdeen University) & Pieter Hemsley (Independent Aviation Consultant)

Important Note

The CAA has made many of the documents that it publishes available electronically (in addition to traditional printed format). Where practical, the opportunity has been taken to incorporate a clearer revised appearance to the documents. Any significant changes to the content of this document will be shown in the Explanatory Note. If no such changes are indicated the material contained in this document, although different in appearance to the previously printed version, is unchanged. Further information about these changes and the latest version of documents can be found at www.caa.co.uk.

23 June 2003

ISBN 0 86039 885 4

Published 23 June 2003

Enquiries regarding the content of this publication should be addressed to:
Research Management Department, Safety Regulation Group, Civil Aviation Authority, Aviation House, Gatwick Airport South, West Sussex, RH6 0YR.

The latest version of this document is available in electronic format at www.caa.co.uk, where you may also register for e-mail notification of amendments.

Printed copies and amendment services are available from: Documedia Solutions Ltd., 37 Windsor Street, Cheltenham, Glos., GL52 2DG.

Check List of Pages

Section	Appendix	Page	Date
		iii	23 June 2003
		iv	23 June 2003
		v	23 June 2003
		vi	23 June 2003
		vii	23 June 2003
		viii	23 June 2003
		ix	23 June 2003
Section 1		1	23 June 2003
Section 1		2	23 June 2003
Section 1		3	23 June 2003
Section 1		4	23 June 2003
Section 1		5	23 June 2003
Section 1		6	23 June 2003
Section 1		7	23 June 2003
Section 1		8	23 June 2003
Section 1		9	23 June 2003
Section 1		10	23 June 2003
Section 1		11	23 June 2003
Section 1		12	23 June 2003
Section 1		13	23 June 2003
Section 1		14	23 June 2003
Section 1		15	23 June 2003
Section 1		16	23 June 2003
Section 1		17	23 June 2003
Section 1		18	23 June 2003
Section 1		19	23 June 2003
Section 1		20	23 June 2003
Section 1		21	23 June 2003
Section 1		22	23 June 2003
Section 1		23	23 June 2003
Section 1		24	23 June 2003
Section 1	Appendix 1	1	23 June 2003
Section 1	Appendix 1	2	23 June 2003
Section 1	Appendix 1	3	23 June 2003
Section 1	Appendix 1	4	23 June 2003
Section 1	Appendix 1	5	23 June 2003
Section 1	Appendix 1	6	23 June 2003
Section 1	Appendix 1	7	23 June 2003
Section 1	Appendix 1	8	23 June 2003
Section 1	Appendix 1	9	23 June 2003
Section 1	Appendix 1	10	23 June 2003
Section 1	Appendix 1	11	23 June 2003
Section 1	Appendix 1	12	23 June 2003
Section 1	Appendix 1	13	23 June 2003
Section 1	Appendix 1	14	23 June 2003
Section 1	Appendix 1	15	23 June 2003
Section 2		1	23 June 2003
Section 2		2	23 June 2003
Section 2		3	23 June 2003
Section 2		4	23 June 2003
Section 2		5	23 June 2003
Section 2		6	23 June 2003
Section 2		7	23 June 2003
Section 2		8	23 June 2003
Section 2		9	23 June 2003
Section 2		10	23 June 2003
Section 2		11	23 June 2003
Section 2		12	23 June 2003
Section 2		13	23 June 2003
Section 2		14	23 June 2003
Section 2		15	23 June 2003
Section 2		16	23 June 2003
Section 2		17	23 June 2003
Section 2		18	23 June 2003
Section 2		19	23 June 2003
Section 2		20	23 June 2003
Section 2		21	23 June 2003
Section 2		22	23 June 2003
Section 2		23	23 June 2003
Section 2	Appendix 1	1	23 June 2003
Section 2	Appendix 2	1	23 June 2003
Section 2	Appendix 3	1	23 June 2003
Section 2	Appendix 3	2	23 June 2003
Section 2	Appendix 3	3	23 June 2003
Section 2	Appendix 3	4	23 June 2003
Section 2	Appendix 3	5	23 June 2003
Section 2	Appendix 3	6	23 June 2003
Section 2	Appendix 3	7	23 June 2003
Section 2	Appendix 4	1	23 June 2003
Section 2	Appendix 4	2	23 June 2003
Section 2	Appendix 4	3	23 June 2003
Section 2	Appendix 4	4	23 June 2003
Section 2	Appendix 4	5	23 June 2003
Section 2	Appendix 5	1	23 June 2003
Section 2	Appendix 5	2	23 June 2003
Section 2	Appendix 5	3	23 June 2003
Section 2	Appendix 5	4	23 June 2003
Section 2	Appendix 6	1	23 June 2003
Section 2	Appendix 6	2	23 June 2003
Section 2	Appendix 6	3	23 June 2003
Section 2	Appendix 6	4	23 June 2003
Section 2	Appendix 6	5	23 June 2003

Section	Appendix	Page	Date
Section 2	Appendix 7	1	23 June 2003
Section 2	Appendix 7	2	23 June 2003
Section 2	Appendix 7	3	23 June 2003
Section 2	Appendix 7	4	23 June 2003
Section 2	Appendix 7	5	23 June 2003
Section 2	Appendix 8	1	23 June 2003
Section 2	Appendix 9	1	23 June 2003
	References	1	23 June 2003
	References	2	23 June 2003
	References	3	23 June 2003
	References	4	23 June 2003
	References	5	23 June 2003
	References	6	23 June 2003
	References	7	23 June 2003
	References	8	23 June 2003

Section	Appendix	Page	Date

Contents

	Check List of Pages	iii
	Acknowledgements	vii
	Glossary of Terms	viii
Section 1	**Literature Review**	
	Executive Summary	1
	Introduction	3
	Method	7
	Results	8
	Discussion	21
	Conclusion	24
Appendix 1	**CRM Evaluation Papers**	
	Civil aviation (N=23 independent studies)	1
	Military (n=17 independent studies)	8
	Other high reliability industries (n=8 independent studies)	13
Section 2	**Methods used to Evaluate the Effectivness of Flight Crew CRM Training**	
	Executive Summary	1
	Introduction	3
	Method	3
	Results	5
	Discussion	16
	Conclusions	22
	Recommendations	22
Appendix 1	**Cover letter**	
Appendix 2	**Questionnaire instructions**	
Appendix 3	**CAA Interview Schedule**	

Appendix 4 **Questionnaire responses: all respondents**

Appendix 5 **Questionnaire responses: large carriers**

Appendix 6 **Questionnaire responses: medium carriers**

Appendix 7 **Questionnaire responses: small carriers**

Appendix 8 **The NOTECHS Behavioural Markers**

Appendix 9 **University of Texas Behavioural Markers**

References 1

Acknowledgements

The authors gratefully acknowledge the co-operation and assistance provided by all those who took the time to complete the questionnaire and especially those who also participated in the interview. We also wish to acknowledge the invaluable assistance of Magdalena Ietswaart (Aberdeen University), Rob Calvert, Steve Griffin and Fiona Merritt (all UK Civil Aviation Authority).

Glossary of Terms

ACAQ	Aircrew Co-ordination Attitudes Questionnaire (military)
AACE	Aircrew Co-ordination Evaluation checklist (military)
ACRM	Advanced Crew Resource Management (US)
AOC	Air Operator Certificate
ATC	Air Traffic Control
ATCSQ	Air Traffic Control Safety Questionnaire (Eurocontrol)
BASIS	British Airways Safety Information System
CAA	Civil Aviation Authority (UK)
CMAQ	Cockpit Management Attitudes Questionnaire (civil pilots)
CRM	Crew Resource Management
FAA	Federal Aviation Administration (US)
FDQAR	Flight Data Quick Access Recorder
FMAQ	Flight Management Attitudes Questionnaire
GIHRE	Group Interaction in High Risk Environments (research project)
JAA	Joint Aviation Authorities (Europe)
JAR-OPS	Joint Aviation Requirements – Flight operations
LLC	Line/LOS Checklist (University of Texas (UT))
LOE	Line Operational Evaluation
LOFT	Line Oriented Flight Training
LOS	Line Oriented Simulation
LOSA	Line Operations Safety Audit (University of Texas (UT))
MRM	Maintenance Resource Management
NOTECHS	Non-Technical Skills behavioural marker system (civil pilots; Europe)
NPA	Notice of Proposed Amendment (relates to JARs)
NTSB	National Transportation Safety Board (US)
PAOC	Police Air Operator Certificate
PHARE	Pologne-Hongrie: aide à la restructuration économique (Action plan for co-ordinated aid to Poland and Hungary; EU research initiative which has been expanded to support all Eastern European countries)
SESMA	Special Event Search and Master Analysis
TARGETs	Targeted Acceptable Responses to Generated Event or Tasks (military)

TOQ	Technical Operations Questionnaire (maintenance personnel)
Type A licence	AOC holders permitted to carry passengers, cargo and mail on aircraft with 20 or more seats.
Type B licence	AOC holders permitted to carry passengers, cargo and mail on aircraft with 20 seats and/or weighing less than 10 tonnes.

Section 1
Literature Review

Executive Summary

This literature review has been prepared as part of a project commissioned by the Safety Regulation Group of the Civil Aviation Authority (CAA) to identify the methods currently avialable to airlines to assess the effectiveness of their Crew Resource Management (CRM) training programmes. CRM training is designed to improve aviation safety. The aim of this review is to establish current knowledge about the evaluation of CRM training effectiveness. This will enable the identification of the methods currently available for the evaluation of CRM training.

A total of 48 research studies were identified following a literature search. These studies were examined to determine the methods used to evaluate CRM training and the results of the analysis. Kirkpatrick's (1976) evaluation hierarchy was used to provide a framework for considering the effects of a training intervention on an organisation. There are four different levels of evaluation:

Level 1: Reactions. This is an assessment of participants' feelings towards the training course. It is valuable for the trainers in providing information on the perceived relevance of the course and for gaining feedback on where improvements can be made. The studies reviewed tended to use paper based questionnaires, distributed after the training course. In general, those studies, which have included an evaluation of CRM training at this level, have found the responses to be positive.

Level 2: Learning. Evaluations of this type are concerned with whether CRM training has had a positive impact on learning, as demonstrated through an examination of both attitude and knowledge change. Attitudes were generally measured using a paper based questionnaire, and the majority of the 24 studies in which they were examined reported a positive shift in attitudes as a result of CRM training. Knowledge was only assessed in six of the studies examined. The most common method used was a multiple-choice test, which typically showed an increase in CRM related knowledge after participation in the training.

Level 3: Behaviour. The most widely used method for evaluating CRM training at this level is by rating CRM skills, using behavioural marker checklists. If the raters have been properly trained and calibrated, a behavioural marker checklist should provide a reliable technique for assessing a range of non-technical skills. As with the evaluations carried out at levels one and two, most studies found that CRM training led to an improvement in CRM behaviour in the cockpit.

Level 4: Organisation. Finally, attempts can be made to assess the impact of CRM training on the organisation as a whole. In civil aviation, only two of the studies reviewed assessed the effects of CRM training at this level, providing evidence of a decrease in violations and prevention of incidents. However, in military aviation, five studies found that CRM training lead to a decrease in the rate of accidents and incidents.

This literature review identifies the available research on the evaluation of CRM training effectiveness and uses this information to make recommendations about how the effects of CRM training should be assessed. However, despite 20 years use of CRM training, surprisingly little published material is available on the evaluation of training. Nevertheless, from the 48 papers reviewed there is ample evidence to suggest that CRM training results in participants reacting positively to the training and achieving positive changes in their attitude, knowledge and behaviour. From the evidence available it is not possible to be as certain about the influence of the training on the organisation as a whole, and whether it has had the ultimate effect of

increasing safety. This is due to the difficulty in attributing changes at this level solely to CRM training, and because there are few studies that have made a rigorous assessment of the effects of such training on the organisation.

This literature review demonstrates that it is important to track the effects of the training to allow for the identification of topics for recurrent training and to ensure that it continues to improve performance, despite changes in aircraft type and pilot demographics. Furthermore, any evaluation of training should be carried out at as many levels as feasible to ensure that as much information as possible is gained. It is also important that the techniques used to carry out any evaluation are valid, reliable, and, if an observational technique is to be used, the observers have been properly trained.

1 Introduction

1.1 Overview

This literature review has been prepared as part of a CAA project commissioned by the Safety Regulation Group of the UK Civil Aviation Authority to identify the methods currently avialable to air operators to assess the effectiveness of their Crew Resource Management (CRM) training programmes.

1.2 Aim

The aim of this literature review is to present a summary of current research on the evaluation of CRM training effectiveness. This will enable the identification of the methods which have been employed to evaluate CRM training and the findings from the studies.

This report:

- gives an outline of CRM training.
- presents a taxonomy for evaluating training programmes.
- discusses the evaluation methods which have been used to assess the effects of CRM training.
- recommends methods of evaluating CRM training.

1.3 Crew Resource Management training

1.3.1 Background

Safety research has shown that it is human error as opposed to mechanical failure that is a major causal factor in industrial and transportation accidents (FAA, 1998; Hollnagel, 1993; Wagenaar & Groenweg, 1987). Although it should be emphasised that organisational policies and managerial behaviour create a safety climate which influences the risk of operator error (HSE, 1999; Weick, Sutcliffe & Obstfeld, 1999). The aviation industry recognised the significance of human error in accidents almost 30 years ago (Maurino, Reason, Johnston & Lee, 1995) and has been instrumental in the development of human factors training programmes known as CRM, designed to reduce error and increase the effectiveness of flight crews (Wiener, Kanki & Helmreich, 1993).

1.3.2 What is Crew Resource Management training?

Crew Resource Management can be defined as 'using all the available resources- information, equipment, and people- to achieve safe and efficient flight operations' (Lauber, 1984: 20). Similarly, an aviation human factors group (Royal Aeronautical Society, 1996:2) define the objectives of CRM to be:

'To enhance crew and management awareness of human factors which could cause or exacerbate incidents which affect flight safety.

To enhance knowledge of human factors and develop CRM skills and attitudes which when appropriately applied could extricate an aircraft operation from incipient accidents and incidents whether perpetrated by technical or human factors failings.

To use CRM knowledge, skills and attitudes to conduct and manage aircraft operations, and fully integrate these techniques throughout every facet of the organisation culture, so as to prevent the onset of incidents and potential accidents.

To use these skills to integrate commercially efficient aircraft operations with safety.

To improve the working environment for crews and all those associated with aircraft operations.'

From accident analyses, incident reports, and research studies both on the flight deck and in the simulator, training packages have been implemented to 'close the loop' between accidents and training (see Figure 1).

Figure 1 'Figure eight' CRM training model.

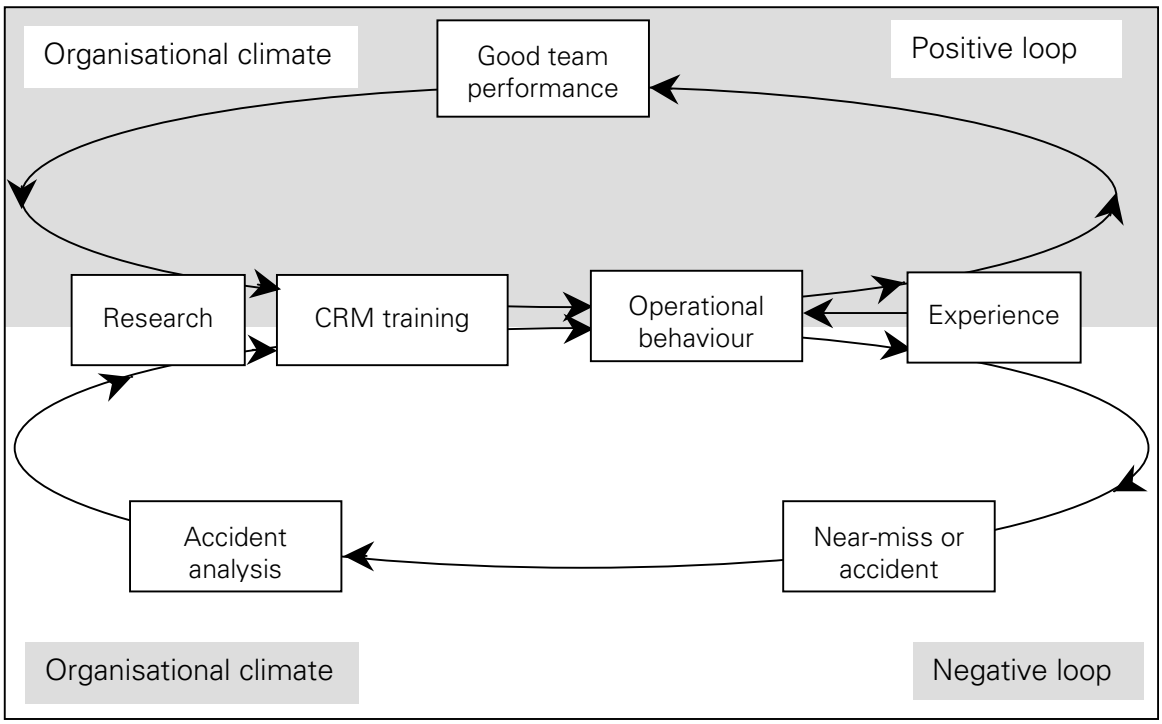


Figure 1 shows that training needs can be identified from both positive and negative input. In the positive loop, where a team performs well, this can influence training by illustrating how good team performance can result in a positive outcome (e.g. United 232; Hayes, 1992). The negative loop shows how failures in team work can be used to identify team training requirements. By carrying out in-depth accident analysis, any trends in poor team working can be identified and be addressed through training. Also, according to the model, teams operate within a particular organisational climate, the effects of which can have either detrimental or beneficial effect on team performance (Hackman, 1983). The kind of norms or values that operate within an organisation, how authority and power are exercised have a powerful effect on the actions and motives of employees (Schein, 1988).

Over the 20 years in which CRM training has been used by the aviation industry, it has focused on: general strategies of interpersonal behaviour; changing attitudes regarding flight deck management; specific aviation concepts relating to group dynamics; the expansion of the training to other personnel, and finally to recognition that human error is inevitable (Salas, Burke, Bowers & Wilson, in press). Currently CRM training is being used as a way of managing these errors by focusing on training which promotes error avoidance and the early detection of errors, and minimises the consequences resulting from human errors (Helmreich, 1996).

CRM training is now used by virtually all the large international airlines and is recommended by the major civil aviation regulators (e.g. FAA, 1998; JAA, 2001). A recent survey of International Air Transport Association affiliated airlines indicated

that 96% of respondents were running CRM courses. Over 60% of these had been in existence for five years or more (O'Leary, 1999). In the UK, human factors training and examination are mandatory for a Flight Crew Licence, and the CAA requires that CRM training be carried out annually by commercial pilots (CAA, 1998a) and, in due course, by cabin crew (CAA, 1995). A typical introductory CRM course takes three days. Teaching methods include lectures, classroom training, practical exercises, case studies and films. The topics covered, 'are designed to target knowledge, skills, and abilities as well as mental attitudes and motives related to cognitive processes and interpersonal relationships' (Gregorich & Wilhelm, 1993: 173). A course typically covers six core topics: team work, leadership, situational awareness, decision making, communication, and personal limitations (Flin & Martin, 1998). Refresher training is also advised. This is normally a half or whole day course focusing on a specific CRM topic. For flight deck crews, CRM skills are then practised and can be assessed in flight simulator sessions known respectively as LOFT (line oriented flight training) and LOE (line operational evaluation). However, it is only recently that European legislation has referred to the evaluation of CRM training, despite its large-scale use in the aviation industry.

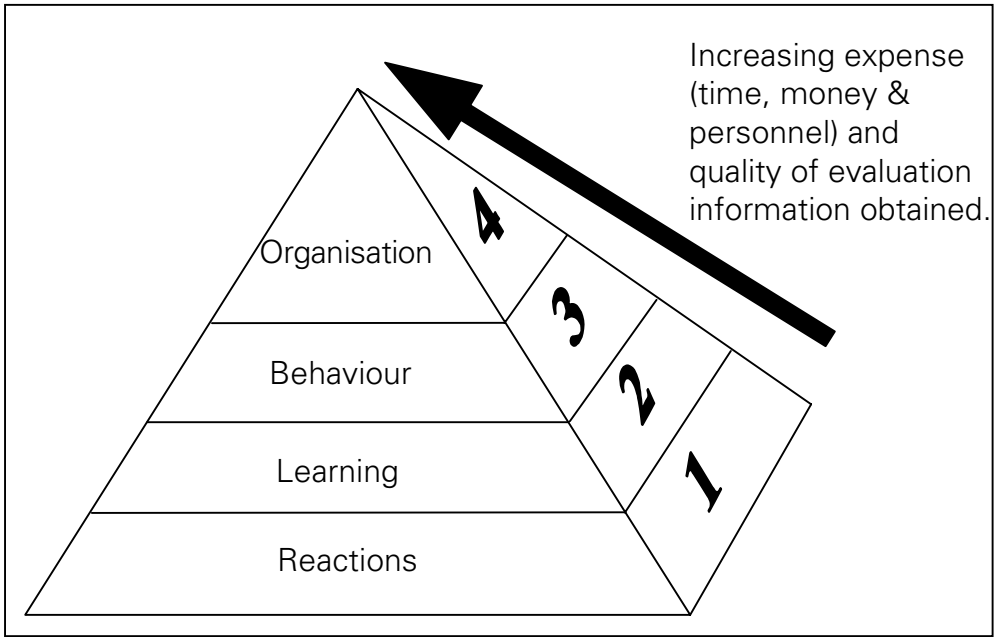
1.4 A framework for evaluating training effectiveness

The CAA has stated that 'the variability of CRM standards and the lack of common practical reference criteria have indicated the need for research into means of assessment' (CAA, 1998b: 1). The Federal Aviation Administration (FAA) also recognises the crucial role of CRM evaluation. 'It is vital that each program be assessed to determine if it is achieving its goals' (FAA, 1993), and further, 'each organisation should have a systematic assessment programme' (FAA, 1998, 14). However, the fundamental question of whether CRM training can fulfil its purposes of increasing safety and efficiency does not have a simple answer (Helmreich, Merritt & Wilhelm, 1999).

There are many anecdotal accounts of accidents that were prevented as a result of utilising CRM skills, and CRM is omnipresent in the aviation industry. However, hard evidence that it is effective is difficult to identify. Salas, Fowlkes, Stout, Milanovich, and Prince (1999) state that, although great advances have been made in team training research, comparatively little research has been conducted to evaluate the effectiveness of particular training strategies, such as CRM.

Salas and Cannon-Bowers (1997) outline a number of principles for evaluating team training which have emerged over time. The recommended approach is one which is multi-faceted and considers several separate methods of assessment. The evaluation methods can be categorised into what is described by training researchers (e.g. Hamblin, 1974; Kirkpatrick, 1976) as different levels of training effects ranging from individual to organisational indicators. Despite previous criticism (e.g. Alliger & Katzman 1997), Kirkpatrick's (1976) hierarchy is still the most popular framework for guiding training evaluation (Salas et al., in press). It provides a useful framework for considering the effects of a training intervention on an organisation by considering training evaluations at four different levels (see Figure 2).

Figure 2 Based on Kirkpatrick's (1976) hierarchy of training evaluation



Kirkpatrick's (1976) hierarchy consists of four different levels of evaluation: reactions, learning, behaviour, and organisation. The four levels represent a sequence of methods to evaluate a training programme. The levels can be thought of as first concentrating on individual effects, followed by team, and then finally organisational change (Holt, Boehm-Davis & Beaubien, 2001). Each level builds on the previous one, with the process becoming more difficult and time-consuming to perform at each higher level, but also providing more valuable information (Kirkpatrick, 1998).

Level 1: Reactions. Reactions are concerned with how the participants react to the training. These are usually recorded in terms of satisfaction or enjoyment using a questionnaire. However, it is important to indicate positive reaction does not ensure learning, although a negative reaction almost certainly reduces the likelihood that this has taken place (Kirkpatrick, 1998).

Level 2: Learning. Learning is the second level in the hierarchy, and refers to 'the principles, facts, and skills which were understood and absorbed by the participants' (Kirkpatrick, 1976: 11). This level is concerned with whether the participant has acquired knowledge or has modified their attitudes or beliefs as a result of attending the training course. It is important to measure learning, as no change in behaviour can be expected if no new knowledge or change in attitudes has occurred.

Level 3: Behaviour. The evaluation at the behaviour level is the assessment of whether knowledge learned in training transfers to actual behaviours on the job or a similar simulated environment. Kirkpatrick (1998) outlines the danger of only carrying out an evaluation at this level of the hierarchy. To illustrate, if no behavioural change was found, an obvious conclusion is that the training was ineffective. Nevertheless, reactions may have been favourable and the learning objectives could have been met. He suggests that a number of conditions must be present to make the jump from a positive evaluation at stages 1 or 2, to positive evaluations at stages 3 and 4. These are:

- The participants must have a desire to change.
- The participants must know what and how to make the change.

- The organisational climate must be conducive to allowing for the change (see earlier).
- The participant must be rewarded for changing (e.g. positive feedback, passing line checks).

Level 4: Organisation. This is the highest level of evaluation in Kirkpatrick's (1976) hierarchy. The ultimate aim of any training programme is to produce tangible evidence at an organisational level, such as an improvement in safety and productivity. The problems with the evaluation of training at this level are that it can be both difficult to establish discernible indicators and to attribute these to the effects of a single training course.

Kirkpatrick's four level hierarchy provides a useful framework for categorising the methods with which researchers have evaluated training. Further, a range of experimental designs can be used to assess the effects of the training at each level. The next section of this report will identify the criteria for studies to be included in the review of techniques used to evaluate CRM training.

2 Method

2.1 Identification of materials

The criterion for the inclusion of a study in this review was that it must include an empirical evaluation of a human factors training course which was designed to improve performance. It was decided to not only concentrate on studies with civil aviation, but also to include military aviation, and other high reliability industries (aviation maintenance, maritime, offshore oil production, and medicine). The rationale for the inclusion of these other industries is that the domain in which the training is being applied is not relevant to this review. Rather, the purpose of the review is to concentrate on how the training course is evaluated. In order to identify CRM evaluation papers, information was drawn from a number of different sources:

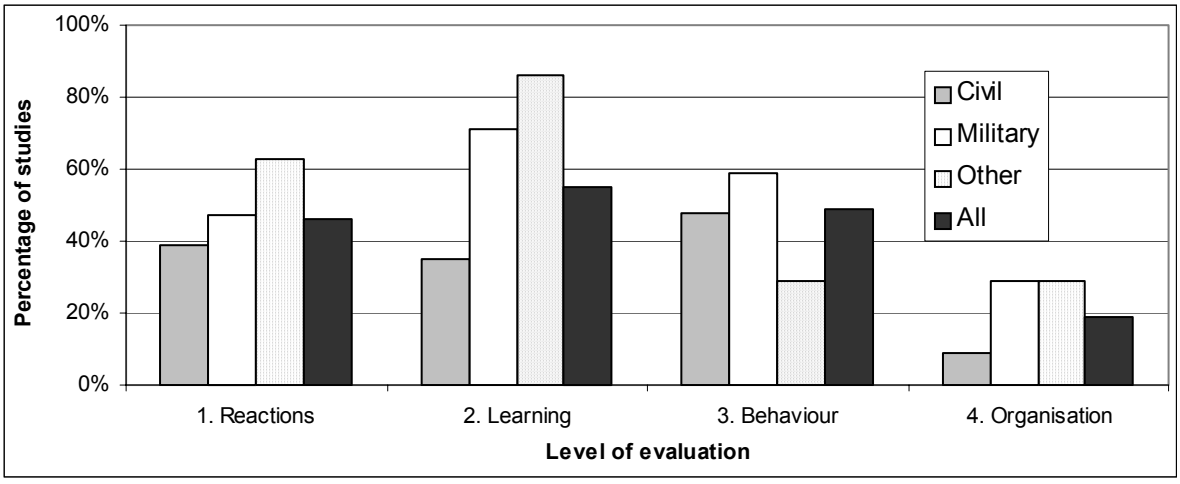
- Aberdeen University's extensive library of research articles.
- Aberdeen University Industrial Psychology Group's previous work with teams in other high reliability industries (offshore, aviation and medicine).
- Material from our contact group of applied psychologists working in aviation training.
- Conventional online data base searches to ensure complete coverage using Web of Science and PsychLIT.
- The contents of conferences papers in which this type of research is published (e.g. International Symposium of Aviation Psychology- vols. 1 to 11-, European Association of Aviation Psychology- vols 1-24) and journals (e.g. Human Factors, International Journal of Aviation Psychology, Journal of Applied Psychology, Journal of Occupational and Organizational Psychology).
- The use of bibliographies of research articles and chapters in relevant papers.
- Contacting authors and research groups known to carry out work in this area.

2.2 Sample

A total of 48 studies were found in which CRM training was evaluated (see Appendix for summary). Of the 48 studies, 23 were from civil aviation (48%), 17 from military aviation (35%), and eight from other high reliability industries (17%; air traffic control, aviation maintenance, offshore oil and gas production, anaesthetics, nuclear power

generation, and the maritime industry). The majority of the studies were carried out by US research teams (36 studies, 75%), with 10 being carried out by European researchers (21%), and two carried out by Australian and Japanese researchers (4%). The sample sizes ranged from 17 to 6354 participants. The studies were categorised by the type or types of evaluation which were reported. This is similar to the technique used by Salas, Burke, Bowers, Wilson (in press) who recently reviewed studies which evaluate CRM training. From Figure 3 it can be seen that the majority of studies examined the effectiveness of the CRM training at the learning level (i.e. attitudes and/or knowledge), with few researchers examining the effectiveness of the training at the organisational level. From an examination of those studies carried out in civil aviation it can be seen that the most common level of evaluation was at the behaviour level, with only two of the studies making an assessment at the level of the organisation.

Figure 3 Percentage of studies carrying out CRM evaluation at each level



In the following sections, the evaluation methods used will be examined in detail, with a discussion of how researchers have carried out evaluations at each of the four levels, and a summary given of the results of the analysis.

3 Results

3.1 Reactions

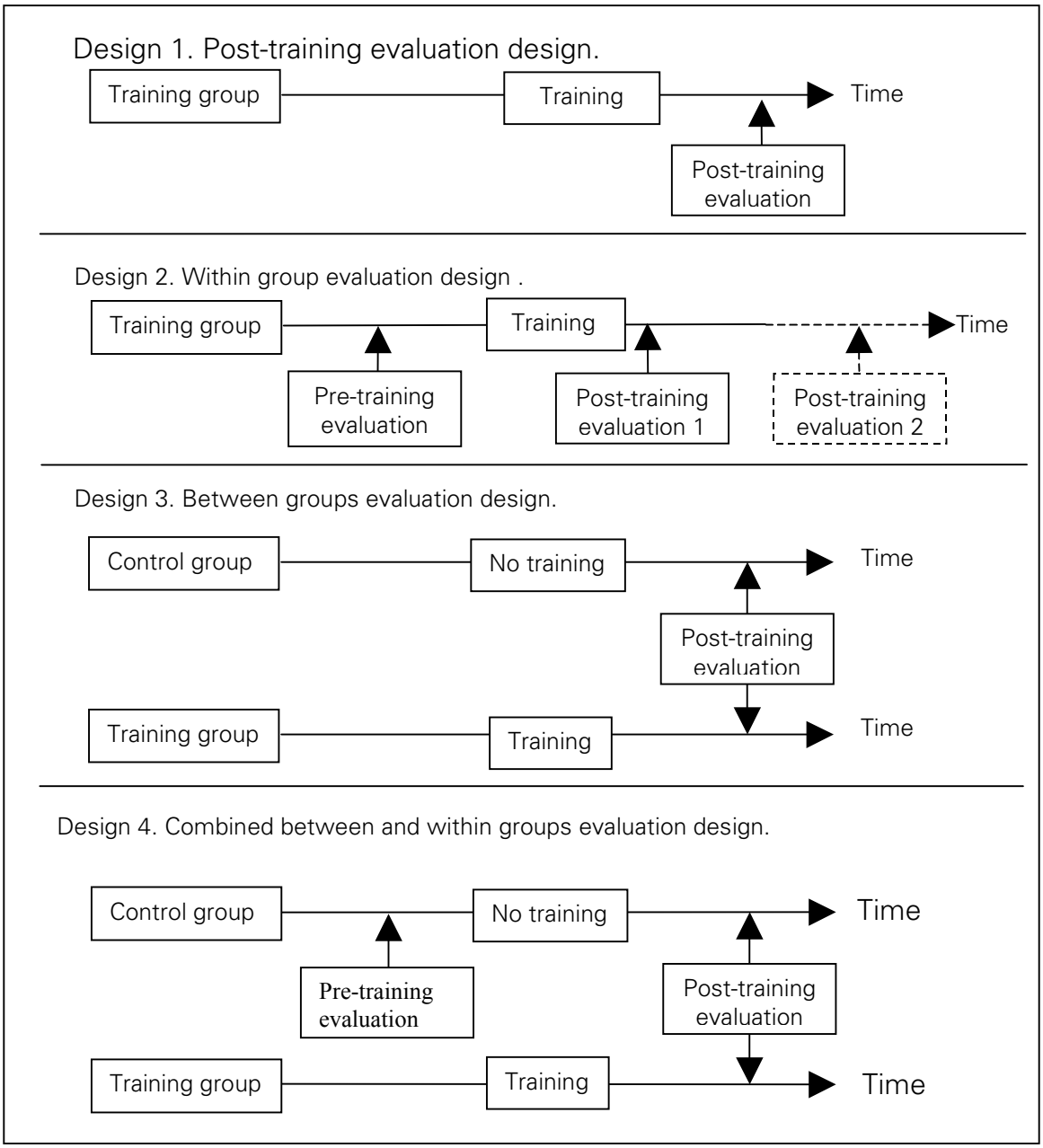
Evaluating reactions is the equivalent to measuring customer satisfaction. This is valuable for the trainers in providing feedback on the relevance of the course and gaining feedback on where improvements can be made. A number of the studies examined carried out an evaluation of participant's reactions: nine studies from civil aviation (39%), eight of the studies from military (47%), and five of the studies carried out in other high reliability industries (63% ; see Figure 3).

A paper based questionnaire method was used exclusively in all of the articles in which the reactions of participants were assessed. Questionnaires are the most common method for obtaining subjective data (Sinclair, 1992). They typically consist of either closed questions for which the respondents can give a fixed range of answers on a response scale (e.g. It was worthwhile attending the course? Do you: strongly agree; agree; neither agree nor disagree; disagree; strongly disagree?).

Questionnaires can also be used to obtain more qualitative information through the use of open questions (e.g. What aspects of the course did you find to be particularly useful?). In all of the studies examined, the questionnaire was distributed at the end of the training course (Post-training evaluation design, see Figure 4).

The studies that measured reactions have generally reported an overall positive evaluation. To illustrate, Taggart and Butler (1989) assessed the reactions of over 2000 flight deck crew members to Pan-Am's Flight Operations Resource Management (FORM) training. Using a paper-based questionnaire it was found that 71% found the seminar to be either very, or extremely useful, and all but 11% indicated that there would be some change in their behaviour on the flight deck. The only study to report some negative reactions was in Schiewe's (1995) assessment of 777 Lufthansa cockpit crew. Its CRM course consisted of a number of different instructional techniques (lectures, group discussion/activity, videos, role play, case studies, individual activity, and peer feedback). It was found that the modules that included training methods using case studies or which allowed participants to act in a job related scenario were rated very positively. However, those modules, which were based exclusively on lectures, were not rated favourably.

Figure 4 Research designs



Summary: Reactions

To summarise, an assessment of participants' reactions is important to establish whether they actually liked the course, and thought that it was useful. This is normally achieved through a post-training questionnaire. This method was used in all of the studies that assessed participants' reactions. In general, evaluations at this level have found the responses to be positive. However, whilst a positive reaction does not ensure learning, a negative reaction almost certainly decreases the chance that any learning has occurred.

3.2 Learning

Evaluation at the level of learning can be carried out to assess the extent to which the participants have acquired knowledge, and/or have modified their attitudes or values

as a result of the training. The methods used to assess attitudes and knowledge are outlined in detail below.

3.2.1 Attitude assessment

The most commonly used experimental design was to assess attitude changes from before and after the training (a within group evaluation design, e.g. Taggart & Butler, 1989; see design 2 in Figure 4). The within group evaluation design was used in 83% of the studies examined. The remaining studies either used a post-training evaluation design (8%, see design 1 in Figure 4) or a between groups evaluation design (8%, see design 3 in Figure 4).

The most common tool for assessing pilots' attitudes to CRM is the Cockpit Management Attitudes Questionnaire (CMAQ; Gregorich & Wilhelm, 1993). This tool was used to assess attitude change in 75% of those studies carried out in civil aviation and 33% of the studies carried out in military aviation which assessed the effects of training at the learning level. The CMAQ is a well-established, training, evaluation and research tool developed to assess the effects of CRM training for flight crew. The CMAQ comprises 25 items chosen to measure a set of attitudes that are either conceptually or empirically related to CRM. The statement topics cover 'Communication and co-ordination', 'Command responsibility', and 'Recognition of stressor effects'. The 'Communication and co-ordination' subscale encompasses communication of intent and plans, delegation of tasks and assignment of responsibilities, and the monitoring of crew members. 'Command responsibility' includes the notion of appropriate leadership and its implications for the delegation of tasks and responsibilities. Disagreement with items on this subscale suggests a belief in the Captain's autocracy. 'Recognition of stressor effects' emphasises the consideration of - and possible compensation for - stressors. Disagreement with items on this subscale suggests a belief in one's own imperviousness to stressors (Chidester, Helmreich, Gregorich & Geis, 1991; Gregorich, Helmreich & Wilhelm, 1990). For each statement in the questionnaire, the degree to which the students agree is assessed using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Where this has been applied before and after CRM training, it allows an assessment of the changes in the attitudes in individual students.

Using the CMAQ, authors have consistently found there to be a positive shift in CRM attitudes as a result of CRM training (e.g. Helmreich, Wilhelm, Kello, Taggart & Butler, 1990; Salas et al., 1999; Helmreich & Wilhelm, 1991). To illustrate, Gregorich et al (1990) found a change in responses towards CRM related attitudes in a survey of 4216 flight deck crew as a result of attending CRM training. Byrnes and Black (1993) used the CMAQ to measure the attitudes of Delta airline pilots and found that CRM training produced a positive attitude change and that this remained stable for up to five years.

However, CRM training does not always lead to a positive attitude change. It has been found that there has always been a small subset of people, known as Boomerangs, Cowboys or Drongos, who have rejected the concept of CRM (Helmreich & Wilhelm, 1991). Efforts at remedial training for these individuals have not been found to be effective. Irwin (1991) found that 8% of a sample of 5830 civilian flight crew in seven different organisations failed to respond to CRM training, and suggested that this may need to be addressed through avenues other than training.

Irwin (1991) also found a decay in positive attitudes to CRM over time. Attitudes were measured at five different time intervals: a baseline (two years prior to initial CRM training); immediately prior to initial CRM training; immediately after initial CRM training; immediately prior to recurrent CRM training (one year later); and immediately after recurrent CRM training. It was found that there was an overall decline in

attitudes on all three CMAQ sub-scales during the intervals between training interventions. The recurrent training resulted in another positive shift in attitudes back to the level found after initial training. Therefore, Irwin (1991) concluded that the reinforcement of CRM concepts through recurrent training is important for attitude maintenance and the stability of attitudes over time. This is also recognised by both the British and US regulators: 'Recurrent training should be implemented now as a natural concomitant to the requirement for initial CRM training' (CAA, 1995: 4.1); 'CRM training must be included as a regular part of the recurrent training requirement' (FAA, 1998:11b). However, Helmreich et al (1999) report a slippage in attitudes to CRM concepts, even with recurrent training.

Proposed possible explanations for the decay in attitudes are the failure of support from management, and failure of evaluators to reinforce the importance of the CRM behaviours in the simulator and on the line. The inclusion of other personnel in CRM training, such as flight attendants, may also have a detrimental effect. This can result in a dilution of the training, which may lack specificity to the pilot's task, and thus reduce the likelihood of an attitude change. Another possible explanation is that CRM courses imported from other cultures, or even different fleets or types of aircraft may have a reduced impact when compared to a course developed for a specific organisational culture and type of operation (Helmreich et al, 1999). Finally, Helmreich et al (1999) also suggest that the basic rationale for CRM training, namely of reducing the frequency and severity of errors that are crew based, may have been lost over time.

The CMAQ provides, at a minimum, a good starting point for the design of an attitude assessment questionnaire. It has been used as the basis of Aircrew Coordination Attitudes Questionnaire (ACAQ) designed for military pilots, the Control Room Operations Attitude Questionnaire designed for nuclear control room personnel (Harrington & Kello, 1992), the Offshore Crew Resource Management Attitude Questionnaire designed for offshore crews (O'Connor & Flin, under review), the Maintenance Resource Management/ Technical Operations Questionnaire (MRM/ TOQ) designed for aviation maintenance personnel (Taylor, 1998), and the Air Traffic Control Safety Questionnaire (ATCSQ; Woldring & Isaac, 1999). The advantage of basing a questionnaire on the CMAQ is that it has been proven to have reasonable psychometric characteristics, and can be used as a good measure to evaluate CRM training programmes (Taylor, 2000b). However, depending on the CRM training, it is possible that the CMAQ will not measure the range of attitudes that should have changed. It could be argued that it does not assess attitudes to issues of decision-making and situation awareness in sufficient detail to provide a useful appraisal of attitudes to these concepts.

Well-designed questionnaires offer the best mechanism for obtaining information on the attitudes of course participants. They can be given to a large number of participants and the responses are relatively easy to collate. However, there are some drawbacks. It is possible that, when assessing any attitude change immediately before and after training, participants may remember the responses which were previously given. It may also be the case that the questionnaire forces participants to give a response on an item to which they have not formed an attitude. This is particularly the case prior to training.

To avoid this problem, particularly when a questionnaire is first being used, an attitude evaluation should not be conducted in isolation. Carrying out evaluations at other levels of the hierarchy can serve as a test of whether similar findings are evident at the different evaluation levels, particularly as the link between attitudes and behaviours is less than perfect (e.g. Abelson, 1972).

3.2.2 **Knowledge assessment.**

In CRM training, as with other training courses, learning can be assessed by testing students on the curriculum. Of the studies reviewed, only seven (15%) report any knowledge assessment. A comparison of knowledge from before and after training (Howard, Gaba, Fish, Yang & Sarnquist, 1992; O'Connor & Flin, under review), or with a control group who have not received the training (Salas, Fowlkes, Stout, Milanovich, Prince, 1999; Stout, Salas & Kraiger, 1996), or a combination of both (Brun, Eid, Johnsen, Ekornås, Laberg & Kobbeltvedt, 2000) provides an indication of what parts of the course have been retained by the participants. Of the papers examined, only Incalcaterra and Holt (1999) assessed any knowledge change in civil aviation pilots. They examined the CRM knowledge gained by 166 Advanced Crew Resource Management (ACRM) trained pilots who were assessed on their knowledge to the new procedures and nomenclature in ACRM training. On the eight question multiple-choice test, seven questions were answered above chance. This was despite the fact that the training had been delivered two years prior to the knowledge assessment. Thus, it was concluded that the ACRM training led to an increase in CRM knowledge.

In military aviation, Salas et al (1999) found that, although CRM training did not show an effect on the pilots' attitudes, it did appear to increase their knowledge of teamwork principles. Those who had participated in the CRM training scored significantly better than the baseline group that had not received any training (a mean of 12.6 out of 17, compared to 9.8 respectively). Stout, Salas, and Kraiger (1996) also attempted to assess knowledge gain with military personnel but found no significant change on a multi-choice knowledge test between the trained and control groups. However, this could be attributed to the very small number of participants (12 trained and 10 controls). Using a multi-choice questionnaire is a quick and simple way of receiving feedback on knowledge acquisition. It can be administered to a large number of individuals and with little effort. However, the questions and answers (both correct and false) must be designed very carefully to avoid either a floor effect (where it is much too hard) or a ceiling effect (where it is much too easy).

Another possible measure of knowledge acquisition is to use written or video scenarios and have the participants attempt to identify the human factors problems and describe how these could have been mitigated. This is arguably a more realistic method of assessing knowledge of human factors as it has greater ecological validity. This was used to assess knowledge after a CRM course delivered to offshore oil platform personnel (see O'Connor & Flin, under review, for more details). The written scenarios were based on real incidents that had occurred offshore and indicated a range of human factors causes. When compared to pre-training responses, it was found that there was a tendency for participants to offer a larger number of possible explanations for the incident after the training, with an increase in the number of explanations classified as situation awareness, decision-making, communication and supervision. However, the use of accident scenarios to evaluate CRM training is a novel method, with no precedent in the literature. Therefore, if this is to be used as a technique to assess CRM training, more work will be required to improve the methodology.

Summary: Learning

The evaluation of training at the learning level is concerned with whether the participant has acquired knowledge or has modified their attitudes and beliefs as a result of the training. It is important to measure learning as no change in behaviour can be expected if there has not been an effect at this level.

Well-designed questionnaires offer the best mechanism for obtaining information on the attitudes of course participants. Attitudes to CRM behaviour in civil aviation (75% of the studies) have generally been measured using the Cockpit Management Attitude Questionnaire (CMAQ). It can be given to a large number of participants and the responses are relatively easy to collate. The advantage of basing an attitude assessment on the effects of CRM training on the CMAQ, is that it has reasonable psychometric characteristics. In general, the CRM training produced a positive attitude change in those studies examined. However, there tends to be a decay in attitudes over time and a minority of pilots appear to retain a negative attitude to CRM behaviour, despite participation in the CRM training.

In those studies examined, knowledge acquisition has been less widely used to assess the effects of CRM training (one study only in civil aviation) than attitude questionnaires. The most common method used in the studies examined was a multiple choice test. Although these are reasonably easy to administer, they need to be carefully designed. In the six studies, which assessed knowledge acquisition in civil aviation, military aviation, anaesthesia and offshore oil crews, five reported an increase in knowledge of CRM skills as a result of the training.

3.3 Behaviour

The third level is the assessment of whether skills learned in training transfers to actual behaviour on the job or a similar simulated environment. A widely used technique for assessing CRM skills in flight crew is for training captains to use an observational and rating system to assess flight crew behaviour. Of the studies reporting an evaluation at this level, five of the studies carried out in civil aviation (45%) and six of the studies carried out in military aviation (60%) employed CRM taxonomy with behavioural markers. These behavioural markers are 'a prescribed set of behaviours indicative of some aspect of performance' (Flin & Martin, 2001: 96). A number of different behavioural marker systems were used in the studies; each of these is discussed below.

3.3.1 Behavioural marker systems

LINE LOS Checklist

The seminal research on behavioural markers comes from Helmreich's group at the University of Texas/NASA/FAA Aerospace Crew research project. In the late 1980s they developed a data collection form called the LINE LOS Checklist (LLC) to gather information on flight crews' CRM performance (Helmreich, Wilhelm, Kello, Taggart & Butler, 1990). The behaviours included in the LLC had their origins in pilot attitudes to cockpit management (Helmreich, 1984) and the analysis of accidents and incidents with identifiable human factors causation (Connelly, 1997). This checklist is widely cited and it has been used as the basis of many airlines' behavioural marker systems (Flin & Martin, 2001). The LLC system has been refined over the years on the basis of ongoing observational research (Clothier, 1991) and more recently was integrated into the Line Operations Safety Audit (LOSA; Helmreich, 2000) instrument. Version

9.0 of LOSA rating form elicits ratings in three broad categories (planning, execution, and review/modify plans) from four phases of flight (pre-departure, take-off and climb, cruise, and approach and landing). Version 9.0 also focuses on threats and errors to the aircraft. Errors committed by the flight crew are described and coded along with actions (if any) taken to mitigate their consequences. As yet, there are no published data using version 9.0 of LOSA to assess the effect of CRM on behaviour. It has been used only as a general human factors auditing tool and has not been used specifically to assess the effects of CRM training. However, earlier versions of the LOSA have been successful in demonstrating the effects of CRM training on pilot behaviour.

Clothier (1991) used a much earlier version (2.0) of the LLC to assess the behaviours of crews on both the line and in LOFT for a US airline before and after CRM training. In this version, ratings are made on a five-point scale of fourteen different behaviours. On the line, a comparison between trained (1000 crews) and untrained (2000 crews) showed that there was a significant difference after training on 12 of the 14 categories. Two areas that showed no significant change were 'concern for accomplishment of tasks' and 'interpersonal relations/group climate'. In LOFT, the 485 trained crews significantly outperformed the 1625 untrained crews in all 14 categories of CRM behaviour. Similarly, Helmreich and Foushee (1991) also used version 2.0 of the Line/LOS checklist to assess the effects of CRM training on pilot behaviour. Over a three-year period of line observations at a major airline, they found a significant positive difference on all 14 categories following the introduction of CRM training.

Version 9.0 of the LOSA should provide an even better tool for assessing CRM training by tracking fleets over time and allowing comparisons to be made between fleets or companies carrying out different types of CRM training. Further, the data on the different error types, and the ability of pilots to mitigate these errors, should provide the information which will allow CRM training to target those areas where it is most required (Helmreich, Klinect & Wilhelm, 2000).

Targeted Acceptable Responses to Generated Events or Tasks (TARGETs)

Another early marker system developed by Fowlkes, Lane, Salas, Franz and Oser (1994) was a team performance measurement approach called Targeted Acceptable Responses to Generated Events or Tasks (TARGETs) designed specifically for military crews. This is based on a set of critical aircrew co-operation behaviours, grouped into seven basic skill areas: mission analysis, adaptability/flexibility, leadership, decision-making, assertiveness, situational awareness and communication. In this system, for each stimulus event in a scenario, there is a predefined set of acceptable behaviours; each is rated as present or absent. As with the LLC, this is a measure of crew performance rather than individual performance. Fowlkes et al (1994) tested the TARGETs approach in a training and evaluation study of six military aircrews and found the measure to have sensitivity and an acceptable degree of inter-rater reliability. Salas et al (1999) used the TARGETs approach to assess US Navy helicopter aircrew and pilots. It was found that the CRM trained crew performed 15% better than the untrained crew during the pre-flight brief and 9% better during high workload segments.

Line Operational Evaluations (LOE) worksheet

Holt and his colleagues at George Mason University have employed a similar, event-based philosophy to assess CRM behaviours as Fowlkes et al's (1994) TARGETs system. Rather than having a generic system, they developed specialised marker systems designed for very specific purposes. For instance, they used a behavioural marker system that was designed specifically to assess a particular situation or scenario. 'The evaluation form emphasises specific crew reactions for these events,

including both technical and CRM performance and related skills' (Holt et al., 2001). Ikomi et al (1999) developed an observation form designed to record the occurrence and effect of proceduralised Advanced Crew Resource Management (ACRM) behaviours in normal operations. By observing 50 line flights in a US regional airline, it was found that those crews who had received ACRM behaviour training showed superior performance on 13 out of the 20 items evaluated. Holt, Boehm-Davis and Hansberger (1999) also used a LOE worksheet designed for a specific scenario. Prior to data analysis, 12 CRM-relevant items were selected to analyse for possible effects of ACRM training. It was found that an ACRM fleet outperformed a non-ACRM fleet in the same airline in six out of the 12 relevant line check items (brief relevant conditions, brief bottom line, communication of decisions, brief airport conditions, brief arrival bottom lines, and use of available resources).

Aircrew Coordination Evaluation checklist

The Aircrew Coordination Evaluation (ACE) checklist was developed to assess military aviators (Leedom & Simon, 1995). Evaluation of team performance is organised around a set of 13 dimensions (e.g. establish and maintain flight team leadership and crew climate, planning and rehearsal accomplishment). These dimensions are also described in terms of specific, operationally relevant, team-related behaviours. Using the ACE checklist, Leedom and Simon (1995) found that after a week of CRM training, US military helicopter crews showed a significant improvement on 12 of the 13 team co-ordination dimensions. They displayed improved communication patterns within the cockpit, more efficient management of crew resource for critical flight tasks, and fewer team errors of the type previously implicated in aviation accidents.

Non-Technical Skills (NOTECHS)

The NOTECHS system is a taxonomy of pilots' non-technical (CRM) skills developed and tested by a consortium of European research organisations and airlines (see Avermaete & Krujsen, 1998). It is divided into four Categories, two of social skills (co-operation, leadership and management) and two of cognitive skills (situation awareness, and decision-making). Each Category is then further subdivided into three or four Elements. In addition, for each Element a number of positive and negative exemplar behaviours are included. The exemplar behaviours are phrased as generic (e.g. closes loop for communication), rather than specific activities (e.g. reads back to ATC; see Avermaete & Kruijsen, 1998 for more details).

The NOTECHS system has been tested using eight video scenarios filmed in a Boeing 757 simulator, which, following a short training session in its use, were rated by 105 instructors from across Europe. The system was well received by the Instructors and was found to be a useful assessment method for both the captain and first officer at the category level (O'Connor et al., in press). The NOTECHS system is currently being tested operationally with instructors using it to evaluate pilots either in the simulator or on standard route flights. NOTECHS is also currently being compared with LOSA as part of a new European project (Group Interaction in High Risk Environments, GIHRE; Häusler, 2000). The aim is to assess which is better able to differentiate between good and poor aircrew performance under high workload conditions.

The NOTECHS system has so far only been used once to assess the effectiveness of CRM training as part of the PHARE Air Safety Improvement Project (ASI; Goeters, 2000). A total of 17 aircrew from an Eastern European airline that had not introduced human factors training participated in the study. When a comparison was made between LOFT performance before and after the CRM training, it was found that there was an improvement in all four NOTECHS categories. This difference was found to be statistically significant for 'situation awareness' and 'decision-making'. The training was not a standard CRM course, but specifically designed to address the

areas of weakness identified in the first LOFT session. Thus, the CRM training was able to shape the professional behaviour in the manner intended (Goeters, 2000).

NOTECHS was recommended by Flin and Martin (1998) following their review of the marker systems being used by aviation companies. Smaller aviation companies in Europe may not have the time, resources or expertise to develop their own behavioural marker systems. This recommendation was also given more weight by a change in the JAA regulations that states: 'the flight crew must be assessed on their CRM skills' (JAA-OPS 1, Change 3, 2001). Therefore, it was recognised that there was a need for a system to assess pilots' CRM skills.

Behavioural marker schemes have been shown to be potentially a very good method for gaining an overview of a flight deck crew's performance. All of the studies reviewed reported an improvement in flight deck crew behaviour as a result of participating in CRM training. However, for these systems to be used effectively for CRM evaluation purposes, it is necessary to ensure that the system is valid and reliable and that the evaluators have been trained to use the system to an acceptable level of reliability (Holt et al, 2001; Klampfer, Flin, Helmreich, Häusler et al, 2001).

3.3.2 Other evaluation methods

A number of other methods were used to assess the extent to which CRM behaviours had changed. These methods generally consisted of responses to questions about whether the course participants would change or had changed their behaviour as a result of the training. To illustrate, Naef (1995) reported that in a follow up survey six months after a nine-day CRM training course, 97% of the flight deck crew reported one or more positive behavioural changes. Another method used was to receive feedback from other team members regarding the behaviour of aircrew who had attended the training. Byrnes and Black (1993) found that, after pilots had attended a CRM course, the flight attendants reported that they were treated with respect, felt more like part of a crew and were more frequently included in crew briefings. Other researchers have concentrated on one particular component of team behaviour. Jentsch, Bowers and Holmes (1995) examined the communication patterns of student pilots using a PC-based flight simulator and found that the CRM-trained crews tended to have more efficient communication patterns compared to the untrained crews.

Simply asking people to give a subjective assessment of whether they have changed their behaviour is not as rigorous as using behavioural markers. The problem with relying on subjective assessment is that the individual will be affected by whether they enjoyed the training or by how well they felt they performed (Venturino, Hamilton & Dvorchak, 1990). Also, the participants will only describe what can be verbalised, which results in information that cannot be verbalised, or is difficult to verbalise, being ignored. Therefore, if the raters have been properly trained, a behavioural marker checklist provides a reliable technique for assessing a range of non-technical skills.

Summary: Behaviour

The evaluation at the behaviour level is the assessment of whether skills learned in training transfer to actual behaviours on the job or a similar simulated environment. The evidence suggests that CRM training can lead to a change in behaviour. The most widely used method of assessing CRM training is by using a behavioural markers checklist. This is superior to self-report techniques of measuring behaviour, such as questionnaires or interviews.

A number of different behavioural marker systems were identified in the studies reviewed, all of which showed an improvement in pilots' CRM performance as a result of CRM training. If the raters have been properly trained and calibrated, then a behavioural marker system applies a structure to their judgements which enhances the objectivity and reliability of their observations. This can allow differences between different fleets or airlines to be identified. Further the data should provide information which will allow CRM training to target those areas where it is most required.

3.4 Organisational Effects

At this level, attempts are made to assess the impact of CRM training on the organisation as a whole. Evidence of an effect at the organisational level is the most valuable evidence of the impact of CRM training. However, only eight of the studies identified in the literature review (17%) carried out any evaluation at this level, with only two from civil aviation. The design used in these studies was a 'within group evaluation design' (Figure 4, design 2). The low number of studies carrying out an evaluation at this level reflects the difficulty in obtaining this type of information. Additional complications are that any change at an organisational level is very difficult to attribute to only one source, given the wide range of other factors which can have an influence (e.g. changes in regulations, organisational structure, aircraft). Therefore, it is necessary to examine a range of measures to assess the impact of CRM training at the organisational level.

In military aviation, where there is a higher accident rate than for civil aviation, five of the studies examined the effect of CRM training at the organisational level. Diehl (1991a, 1991b) found that CRM training reduced aircrew error accident rates by as much as 81 percent in the US military. Alkov (1989, 1991) found that CRM training reduced mishaps due to aircrew error for US naval aircrews. Over a 4-year period there was a reduction in error rates for helicopter pilots from 7.01 (per 100,000 flight hours) to 5.05; for bomber crews a reduction from 7.56 to 1.43; and multi-crew fighter aircraft a reduction from 13.78 to 6.27. Furthermore, the cost of the five-year CRM programme was less than a million dollars, which represents a large financial saving when considering the cost of aircraft and human lives saved. Grubb et al (2001) attributed a reduction in accident rates from 1.75 (1993) to 0.75 (1996) per 100,000 flight hours in US Army aviation to Aircrew Co-ordination Training. Similarly, in the maritime industry, Byrdorf (1998), found that incidents and accidents in the *Maersk* shipping company decreased by a third from one major accident per 30 ship years in 1992 (before the introduction of CRM training) to one major accident per 90 ship years in 1996 (after the introduction of CRM training). Furthermore, at the beginning of 1998 insurance premiums were lowered by 15 percent. This reduction in accidents and incidents was attributed to the CRM and simulator training.

In civil aviation, the accident rate is so low that it does not provide a robust test for the effectiveness of CRM training. Of the studies reviewed, only two examples of CRM evaluation at the organisational level were from civil aviation. Kayten (1993) cites several examples of NTSB reports in which good CRM practices were reported to

limit the effects of either human or mechanical failures. Byrnes and Black (1993) found that the CRM programme at Delta Airlines resulted in a reduction in the 'quarterly air carrier discrepancy reports', which are reports of incidents passed to the airline's operations department (Beneigh, Embry-Riddle University, by correspondence).

Although not used in the studies examined, suggestions have been made of metrics which could be used to assess the effectiveness of CRM training at the organisational level. The Royal Aeronautical Society Human Factors Group (1999) identifies a number of potential measures for assessing the effects of CRM training. These include fuel management, punctuality, job satisfaction, insurance costs and damage to aircraft. Another possible assessment method is to use an organisational climate tool. Boehm-Davis, Holt and Seamster (2001) suggest that this is a useful process to carry out prior to the introduction of CRM training in an organisation to help identify training requirements. However, it is important that the results of such an evaluation method are examined with caution because other changes in the organisation (e.g. down sizing or structural changes) could also affect the responses to a survey.

Summary: Organisational effects

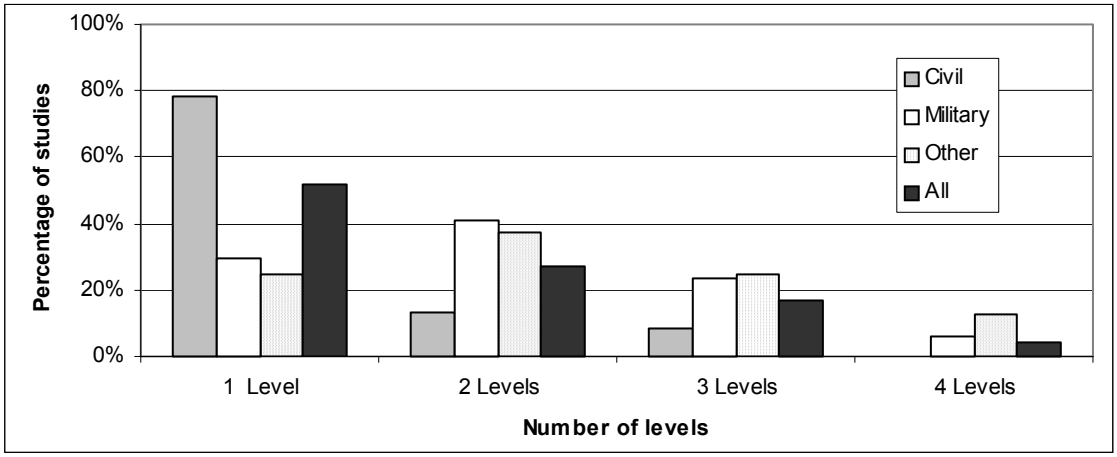
The ultimate aim of a CRM training programme is to produce tangible evidence of an effect at an organisational level such as an improvement in safety and productivity. However, it would appear that, at present, evaluations of CRM training carried out at the organisational level are infrequent. The difficulty of evaluating training at this level is that it can be both difficult to establish discernible indicators and to be able to attribute these to the effects of a single training course. This is particularly the case in civil aviation for which there are few accidents. However, in military aviation and merchant shipping, in which accidents and incidents are more common, the studies examined attributed a reduction in the number of accidents and incidents to CRM training.

In civil aviation it may be of benefit to use additional metrics, such as fuel management or insurance costs as well as the accident or incident rates, to assess the effect of CRM training.

3.5 Multi-level analysis

The best approach used by CRM research teams in aviation is one that is multifaceted and considers several separate methods of assessment (Kraiger, 1993). However, from Figure 5 it can be seen that this was not generally found to be utilised in the studies examined. Particularly in the case of studies carried out in civil aviation, researchers tended to undertake an evaluation at only one of the levels. In civil aviation, a mean of 1.3 levels were evaluated per study, 2.1 in military aviation and 2.1 in other high reliability industries.

Figure 5 Number of levels of evaluation performed by the studies.



The only studies which looked at the effectiveness of the training at all four levels of the hierarchy were Grubb et al. (2001) in military aviation and Taylor (1998) with aviation maintenance personnel. Using a multi-level analysis approach allows the return on investment (ROI) of the training to be calculated. Taylor (2000a) proposes the following equation for calculating the ROI of CRM training.

$$ROI = \left(\frac{[\text{Net CRM Benefits}] \times \text{Causal Operator}}{\text{CRM costs}}\right) \times 100$$

The net CRM benefits are the benefits of the training minus the cost of the development of the training (CRM costs). These are financially calculable savings such as reduction in accidents, reduction in damage to aircraft, increased productivity, etc. The causal operator is a corrective calculation which is used to take into account the many different factors which may act to improve the behaviours or outcomes. The value of the causal operator is the square of the correlation between the CRM training outcome (i.e. knowledge gained, improvement in attitudes or behaviours) and subsequent safety results. The square of a given correlation coefficient is equal to the proportion of the variance in one of the two measures explained by the other and is known as the coefficient of determination (Howell, 1992). This provides 'conservative, quantitative estimates of 'credit' to be allocated to organisational effectiveness (e.g. Maintenance Resource Management (MRM)) interventions' (Taylor, 2000a: 4).

Taylor (2000a) gives the following example to illustrate how this equation is used. The development of an MRM course and delivery to 1,600 employees costs $251,660. A post-training survey showed a significant improvement in attitudes, and in the two years following the training there was a decrease in lost time injuries (LTI) of 80%. A correlation of -.24 was found between LTI and attitudes towards participative leadership and assertiveness in a post-training survey. Therefore, the coefficient of determination is $.24^2$ or 0.0576. Taylor (2000a) estimated the cost of an injury as $13,465. Thus, a reduction of 80% on 91 incidents per year is a saving of $1,314,150. Therefore, the ROI calculation is as follows:

$$ROI = \left(\frac{[\$1,314,150 - \$251,660] \times .0576}{\$251,660}\right) \times 100 = 24.3\%$$

It follows that even with the conservative estimate of 5.76% LTI benefit accounted for by the MRM training, the training paid for itself plus an additional 24% return in two years.

Possible reasons for the lack of multi-level analysis are the costs associated with carrying out the evaluation, the operational constraints of the aviation industry (Salas et al., 1999) and the lack of expertise in performing certain types of analysis. An

explanation for the examination of a greater number of levels in the military than in civil aviation is that the studies tended to be carried out by specialist research groups who had the time, money and expertise to conduct a more in-depth analysis.

There are distinct limitations with carrying out an evaluation at only one or two levels of the hierarchy. It can result in a restricted assessment of the impact of the training. As described earlier, Kirkpatrick (1998) outlines the danger of carrying out an evaluation at only the reaction level of the hierarchy. It is possible that the training course could receive very good feedback. However, this does not necessarily mean that the participants gained any knowledge or changed their attitudes or behaviour.

An in-depth evaluation of effectiveness can help shape the direction of recurrent training by identifying areas of strength and weakness and allow differences to be identified between different fleets of aircraft. As aircraft become increasingly automated, it is critical that operators carry out evaluations of their CRM training courses to ensure that they continue to have a positive effect on pilots' knowledge, skills and attitudes.

4 Discussion

4.1 General summary

This report has used a four level evaluation hierarchy (Kirkpatrick, 1976) to classify the methods and discuss the results of 48 studies which have been used to assess the effects of CRM training. The lack of a large literature on the evaluation of CRM training is surprising. As CRM training is now used by virtually all the large international airlines and is recommended by the major civil aviation regulators (e.g. FAA, 1998; JAA, 2001), more evidence demonstrating its effectiveness was expected than was found. Nevertheless, it is recognised that there may be company specific evaluations of CRM training that have not been published.

The 48 studies examined showed that in general CRM training was well received by pilots, resulted in a positive change in their CRM attitudes and had the desired effect on their CRM behaviours. Only six of the studies reviewed assessed knowledge and only one of these was carried out in civil aviation and little evidence was found of evaluations carried out at the organisational level. Hence it is difficult to draw firm conclusions on the effects of CRM training on knowledge and the organisation. Those studies which did examine effectiveness at this level were carried out primarily using military flight crew and were based on accident and incident frequencies. Thus, there is a need to identify other metrics to allow further evaluation of the effects of CRM training at an organisational level to be made.

4.2 Experimental design

Ideally, CRM training should be assessed at more than one level of the hierarchy. From the literature, it is possible to make a number of recommendations about how CRM training should be evaluated at each level (Table 1).

Table 1 Recommended methods and design for evaluation at each level

Level	Method	Design
1. Reactions	Paper based questionnaire	Post-training (Figure 4, design 1)
2. Learning	Attitude: Attitude questionnaire (e.g. based on CMAQ).	Within group (Figure 4, design 2) Between & within (Figure 4, design 4)
	Knowledge: Written knowledge test	Within group (Figure 4, design 2)
3. Behaviour	Behavioural marker checklist, e.g. NOTECHS, LOSA	Within group (Figure 4, design 2) Between & within (Figure 4, design 4)
4. Organisation	Accident and near-miss data Other performance measures, e.g. fuel management, punctuality.	Within group (Figure 4, design 2) Between & within (Figure 4, design 4)

4.3 Reaction

At the course feedback stage, a paper-based questionnaire is sufficient to gain feedback on the participants' reactions to the training. This questionnaire should be administered after each module of the course has been completed (Figure 4, design 1). Further, the questions pertaining to each course module should be completed after the module is finished, rather than waiting until the end of the course. After two days of training, it may be difficult for the participants to remember the comment they wished to make on the earlier modules. It is suggested that the questionnaire consists of closed statements in which the participants could respond on a five point Likert scale from 1 (very poor) to 5 (excellent) and with open-ended questions allowing them to write their comments.

4.3.1 Learning

At the learning level, attitude change and knowledge can be assessed. The attitude assessment can be carried out using a paper-based attitude questionnaire, which should, at the minimum, be based on an established instrument such as the CMAQ (Table 1). It is suggested that this is completed prior to the training course to obtain a baseline, immediately after the training, and then - if practical - on at least one occasion after a delay of approximately six months to assess the extent of any decay in attitudes (Figure 4, design 2). Ideally, the baseline measure should be taken a couple of months prior to the training, rather than at the start of the course. This will help eliminate any memory effect on the immediately post-training attitude assessment. It may also be useful to compare the responses to a control group who did not receive the training. However, this should be carried out in addition to the within group design (Figure 4, design 4), rather than as an alternative (Figure 4, design 3).

Although not widely used in the studies reviewed, knowledge assessment is a useful method for evaluating the effects of CRM training. The suggested method is a paper-based test (Table 1). This is a fairly quick and simple way of receiving feedback on knowledge acquisition. However, the questions and answers (both correct and false, if using a multi-choice test) must be designed very carefully to avoid either a floor effect (the test is so difficult that the majority of individuals receive very low scores) or a ceiling effect (the test is so easy that the majority of individuals receive very high scores). There are problems with establishing an ideal design for carrying out this type of evaluation. The questionnaire could be administered before and after the training (Figure 4, design 2). However, if the same questionnaire was to be used, there may be a practice effect. Hence, it is desirable to use two different questionnaires of

comparable difficulty. To address this issue, the questionnaire could be completed on two occasions by a control group who have not received the training and at a time interval equivalent to that of a group undergoing training. If no significant difference in results is found, it can be concluded that there is no practice effect. It is recognised that the control group will not be a group of flight crew who have never received CRM training. Rather, if a new type of training is being assessed, flight crew who have yet to undertake the training could serve as a control. This technique was used by Holt et al (1999), Ikomi et al (1999) and Incalcaterra and Holt (1999) who compared CRM (control group) and ACRM (experimental group) trained pilots.

4.3.2 **Behaviour**

To carry out an evaluation of behavioural change, it is suggested that a behavioural marker system, such as NOTECHS or LOSA, is used (see Table 1). Also, many companies already have their own behavioural marker system (Flin & Martin, 2001). However, no matter which system is used, it must have been assessed for reliability and validity. It is also crucial that those instructor pilots who are carrying out the evaluation have been properly trained and calibrated. 'Instructors and check pilots require special training in order to calibrate and standardise their own skills' (FAA, 1993).

Baker and Mulqueen (1999) state that rater calibration training is the strategy which has been advocated for training the pilots who will carry out the observation to ensure accurate and reliable LOE. Typically the training consists of a one-day workshop, in which information about the rating system and practice ratings of videotaped performances of crews flying LOE scenarios are carried out. The participant's ratings are compared to allow the identification of significant discrepancies. These differences are discussed and then a videotape of a different crew flying the same LOE is rated to determine the level of calibration achieved (George Mason University, 1996; Williams, Holt & Boehm-Davies, 1997). However, only a few studies have evaluated the effectiveness of rater calibration training and the results have been mixed (Baker and Mulqueen, 1999). To illustrate, Williams et al (1997) reported acceptable levels of rater agreement, but rater calibration training did not result in a significant improvement. Therefore, Baker and Mulqueen (1999) suggest that raters are also trained in observational skills and practise and receive feedback on the rating task.

Ideally, a pilot's behaviour should be assessed both on the line and in the simulator. On the line, an audit can be carried out before and after the training. Then, if possible, a further audit should be undertaken on at least one occasion after a delay to assess the extent of any decay in behaviours (see Figure 4, design 2). Comparisons could also be made with a control group. However, as with an attitude assessment, this method should be carried out in addition to the within group design (Figure 4, design 4) rather than as an alternative (Figure 4, design 3).

4.3.3 **Organisation**

Identifying a method for evaluating CRM training at the organisational level is the greatest challenge but, as yet, there are no established reliable measures which have been used in civil aviation. Once reliable measures have been established this method of assessment should be carried out prior to training to establish a baseline and then repeated on at least one occasion after training (Figure 4, design 2). In addition, it is also theoretically possible to make comparisons with a control group who had not received the training (Figure 4, design 4).

The experimental designs identified in Figure 4 should enable valid and reliable methods of assessing CRM training to be developed. The evaluation of CRM training

offers a means of tracking the continued effectiveness of CRM training in improving aviation safety and of assessing the impact of any new methods of training.

5 Conclusion

This literature review has identified the available research on the evaluation of CRM training effectiveness and used this information to make recommendations about how the effects of CRM training should be assessed. However, despite over 20 years of the use of CRM training, surprisingly little published material is available on the evaluation of training. Nevertheless, from the 48 papers reviewed, there is ample evidence to suggest that CRM training results in participants reacting positively to the training and achieving positive changes in their attitude, knowledge and behaviour. From the evidence available, it is not possible to be as certain about the influence of the training on the organisation as a whole and whether it has had the ultimate effect of increasing safety. This is due to the difficulty in attributing any changes at this level solely to CRM training and because there are few studies which have made a rigorous assessment of the effects of such training on the organisation. Nevertheless, Taylor's (2000) return on investment measure offers a mechanism for demonstrating the effectiveness of the CRM training that then allows comparisons to be made between organisations, training programmes, industries, etc.

The literature review has demonstrated that it is important to track the effects of CRM training to allow for the identification of topics for recurrent programmes and to ensure that it continues to improve performance despite changes in aircraft design, operational conditions, emerging risks and pilot demographics. As Gregorich and Wilhelm (1993) have argued, any evaluation should be carried out at multiple levels, with assessment techniques of proven validity and reliability. Proper evaluation data could be used for internal performance auditing, as well as for benchmarking across companies and industrial sectors to ensure an optimal return on CRM training investment.

Appendix 1 CRM Evaluation Papers

1 Civil aviation (N=23 independent studies)

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Butler (1993), USA.	Cross fleet/cross airline. Number of participants not specified	Initial CRM training, no details of topics provided.	Most crews evaluated CRM favourably.			
Byrnes & Black (1993), USA.	Delta Airlines crews. Number of participants not specified	8 hour, 4 hour session. Topics covered included: Traditions Foundations Crew dynamics Communication Decision dynamics Stress management		Significant positive change in attitudes as measured by the CMAQ.	Anecdotal evidence from flight attendants that they were treated with more respect, felt more a part of a crew, and included more often in crew briefings.	Quarterly air carrier discrepancy reports declined, FAA violation cases declined.
Clark, Nielsen & Wood, (1991), USA.	135 commercial airline pilots.	Initial CRM training, no details of topics provided.	From a survey it was concluded that the majority of pilots sampled experienced enhanced information processing as a result of the training and the majority of the samples indicated that the negative effects of aviation stress outweighed the positive effect of CRM training.			

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Clothier (1991), USA.	Major US domestic airline (3000 crews; 2000 untrained 1000 trained)	Initial CRM training, no details of topics provided.			On the line, the trained crews out performed untrained crews in 12/14 LLC categories. Significant improvement in LOFT exercises measured by the LLC (485 trained out performed 1625 crews on all 14 categories)	
Goeters (2000), Germany.	17 Eastern European pilots from a National Airline	Lufthansa CRM training. Topics covered included: Management Communication Teamwork Decision making			Using the NOTECHS system, crews improved in all aspects of CRM skills pre vs. post CRM training on LOFT.	
Gregorich et al. (1990), USA.	US National Air carrier (n= 380) cockpit crew Major air carrier (n= 3836) cockpit crew	Initial CRM training, no details of topics provided.		Positive attitude change as reflected in responses to CMAQ Reduction in variation for Communication and co-ordination and command responsibility		
Gregorich (1993), USA.	1191 flight crew from a major air carrier.	Initial CRM training, no details of topics provided.		As measured using the CMAQ, initial training led to a positive change in attitudes. However, there was a significant reduction in attitude levels between training cycles.		

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Hansberger, Holt & Boehm-Davies (1999), USA.	19 Inspector/ Evaluators (I/E); considering all pilots they evaluated over a 6 month period	Topics covered included: Communication Situation assessment Planning/decision making			Using a 19 item fleet comparison questionnaire the ACRM pilots were assessed higher than non ACRM pilots on workload management, communication, and planning.	
Helmreich & Foushee (1993), USA.	Major US airline. Number of participants not specified.	Initial CRM training, no details of topics provided.			Significant positive shift towards effective CRM behaviour during line operations across a 3 year period as measured by the Line/ LOS checklist.	
Holt, Boehm-Davis & Hansberger (1999), USA.	All pilots in ACRM fleet and regional carrier (control; who only received CRM training).	Topics covered include Communication Situation assessment Planning and decision making.			As assessed in LOE using a LOE worksheet, the ACRM fleet performed better than the non-ACRM fleet on comparable LOE items. The ACRM fleet performed better than the control on 6 from 12 CRM-related line check items.	

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Ikomi, Boehm-Davis, Holt & Incalcaterra (1999), USA.	50 crews from an Eastern US regional airline. Comparison of two fleets (ACRM vs. non ACRM).	Topics covered included: Communication Situation assessment Planning/decision making			Performance compared on a 20 item LOE worksheet, ACRM pilots showed superior performance in 13/20 items.	
Incalcaterra & Holt (1999), USA.	184 ACRM trained, 84 non-ACRM trained pilots.	Topics covered included: Communication Situation assessment Planning/decision making	93% thought ACRM should be expanded to other fleets. 91% felt it had improved their flight performance.	As measured by CMAQ, ACRM trained pilots had positive attitudes to CRM. In the multi-choice knowledge test, performance was better than chance on 7/8 items.		
Irwin (1991), USA.	Flight crew from 7 organisations (US, other carriers, and US military) (5830 pre/post initial CRM; 6354 recurrent CRM after a year.	Initial CRM training, no details of topics provided.		As measured using the CMAQ, there was a significant positive change in attitudes. Few boomerangs, attitudes declined over time, recurrent training led to an increase in positive attitudes.		

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Jentsch, Bowers & Holmes (1995), USA.	20 instrument-rated pilots from Embry-Riddle Aeronautical University.	Topics covered included: Situation awareness Mission analysis			Using an 8 category crew interaction coding form, it was found that the communication ratio decreased for crews receiving training. Crews tended to maintain efficient communication patterns. Communication frequencies did not change significantly after 45 days.	
Kayten (1993), USA	NTSB reports.	Content of training not relevant to study.				Several near-miss incidents saved from catastrophe attributed to CRM training.
Margerison, Davies & McCann (1987), Australia.	Australian airline	Topics covered included: Decision making Conflict resolution Crew reactions Interpersonal understanding	Favourable reactions	As measured by the Margerison/McCann Team Management Index: Aircrew have a better understanding of how they and others do the job Discuss with each other how they prefer to work Greater tolerance and understanding between crew members		

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Maschke, Goeters, Hormann & Schiewe (1995), Germany.	224 Pilots in DLR/ Lufthansa training.	3 day course. Topics covered included: Communication Leadership and teamwork Judgement and decision making Stress management	80% thought the course was extremely useful.			
Naef (1995), Germany.	Swissair line and instructor pilots.	9 day training course over 18-month period. Topics covered include: Communication Feedback Decision making Team working Functioning under pressure Automation	Majority of participants thought the course was well presented. The vast majority of instructors thought the course was worthwhile and related to their practical needs.		In a follow up survey 6 months later, 97% of flight crew reported one or more positive behaviour transfer(s). 70% of instructors would invest an off-duty day per year towards soft skills training.	
Predmore (1991), USA	United Flight 811 and Flight 232.	Content of training not relevant to study.			Both Captains of the two flights cited CRM training as contributing significantly to the performance effectiveness of the crews in these accidents.	

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Schiewe (1995), Germany	Boeing 777 German Lufthansa cockpit members.	3 day CRM training course. Topics included: Communication Judgement & decision making Teamwork	Units that were based on case studies or used role play in job related scenarios were rated very positively Units based mostly on lectures were not rated favourably.			
Taggart & Butler (1989), USA.	>2000 cockpit crew members trained in Pan-Am.	Initial CRM training, no details of topics provided.	Generally, participants thought the training was useful.	Significant positive attitude change as measured by CMAQ.	Improvement in crew performance on the line, measurement method not described.	
Vandermark (1991), USA.	American West Flight attendants and cockpit crew (approx. 1200).	2 days of training. No details of topics provided.	Positive reaction to training			
Yamamori & Mito (1993), Japan.	2300 members of Japan airline.	4 days of training. No details of topics provided.		Strengthens attitudes to CRM concepts as measured by a questionnaire.		

2 Military (n=17 independent studies)

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Alkov (1989), USA.	26 US Navy and Marine helicopter instructor pilots.	LOFT type CRM training. No details of topics provided.		Positive shift in attitudes to CRM as measured by the CMAQ.		Training in squadrons helped to reduce mishaps due to aircrew error from 59.8% to 50%.
Alkov (1991); Alkov & Gaynor (1991), USA; same data used in both studies.	90 naval aircrew members (65 completed before/ after survey).	Aircrew co-ordination training, consists of 2-3 days training. Topics covered included: Pilot judgement Situation awareness Policy & regulations Command authority Workload performance Use of available resources Operating strategies Communication skills		Modified CMAQ showed a positive change in attitudes regarding CRM behaviour.		Aircrew error mishap rate declined.
Baker, Bauman & Zalesny (1991), USA.	41 US Navy CH-53 helicopter pilots.	Training covered: Pre-flight brief Assertiveness	Overall, trainees felt that the pre-flight brief exercise was a valuable addition to the course. Review of mean responses indicated that the exercise was perceived as worthwhile and likely to impact on the way they gave/ received their next briefing Open-ended questions provided evidence of a good impression on both exercises. Positive reactions to the assertiveness exercise.			

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Baker, Prince, Shrestha, Oser & Salas (1993), USA.	112 US Male military aviators.	Initial CRM training, no details of topics provided.	90% of participants agreed that the tabletop systems used in training could be used to augment CRM training.			
Baker, Clothier, Woody, McKinney & Brown (1996), USA.	17 Active duty USAF crews flying tanker aircraft.	Initial CRM training, no details of topics provided. Compared fixed vs. formed crews.			Using the NASA/UT Line LOS checklist. No significant difference between fixed and formed.	
Chidester et al. (1991), USA.	531 USAF military airlift command pilots.	Initial CRM training, no details of topics provided.		As measured by the CMAQ training generally produced positive attitudes, however there was an effect on personality type which was also measured.		
Diehl (1991a); Diehl, (1991b), USA; same data used in both studies.	Military/civil aviation. Numbers not specified.	Initial CRM training, no details of topics provided.			Reduction in errors ranging from 8% to 46% using different training methods.	Reduction in accidents ranging from 28% to 81% decrease.
Dwyer, Fowlkes, Oser, Salas & Lane (1997), USA	19 military participants from close air support.	5 days of training, no details of topics provided.			Overall increase in performance across 5 days of training as measured by TARGETs.	

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Elliott-Mabey (1999), UK	Responses from 3212 UK RAF air crew. Tracked attitudes to CRM in RAF over a 3 year period from 1996 (introduction of CRM) until 1998.	Initial CRM training, no details of topics provided.	Positive support for CRM for all aircraft type, particularly multi-engine.	Using Aircrew Attitude Questionnaire, there was a significant positive change in attitudes (communication and co-ordination, leadership and command, conflict resolution, stress and situation awareness).		
Geis (1987), USA.	Pre-test 838 US Army pilots, post 163. 142 responses from a delayed survey, 3 months after the course.	Topics covered include: Introduction Decision making Communication Conflict resolution Situation awareness Procedural compliance Judgement Problem solving Prioritisation Workload management Stress Distractions	Very positive response to the training course and content.	As measured by the CMAQ there was a significant positive change in attitudes.		
Grubb et al. (2001), USA.	US Army aircrew	Air crew co-ordination training (ACT), no details of topics provided.		Statistically significant improvements when comparing pre/post-training attitudes on 12/13 basic qualities as measured by an attitude survey.	Increased emphasis on the importance of CRM by unit leaders and air crews.	50% reduction in Class A accident rates as a result of an emphasis on Army ACT.

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
	US Airforce aircrew	Initial CRM training, no details of topics provided.	Airforce aircrew consistently reported the positive effect of CRM training on flying safety and mission effectiveness.	Statistically significant improvement in CRM attitudes as measured by a 24 item questionnaire (3600 personnel). However, there was a 25% reduction in attitudes towards CRM after 12-14 months.	Increased emphasis on the importance of CRM by unit leaders and air crews.	Human factors related mishaps at approximately 60% (20% less than US General Accounting estimates) in the Airforce.
Leedom & Simon (1995), USA.	32 military US pilots (UH-60 crews).	1 week of training, no details of topics provided		Small positive shift in attitudes regarding team co-ordination on Army Aviation Crewmember questionnaire (AACQ).	Using the Aircrew Co-ordination Evaluation (ACE) checklist significant improvement in 12 of the 13 dimensions. Improved communication patterns, more efficient management of crew resources during critical phases of flight, fewer errors of the type previously implicated in aviation accidents.	
	30 AH-64 US aviators.	1 week of training, content not specified, no details of topics provided		AACQ should no statistically significant change in attitudes regarding team co-ordination.	Improvement across all 13 dimensions of ACE checklist (statistically significant in 6). Improvement in team co-ordination and performance.	

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Salas et al. (1999), USA.	35 pilots and 34 enlisted aircrew from US Navy transport helicopters.	Topics covered in training included: Introduction Communication Assertiveness Mission analysis Situation awareness	Strong endorsement of the usefulness of the training.	Attitudes measured using the Aircrew Coordination Attitudes Questionnaire (ACAQ; based on CMAQ and AACQ) showed an overall positive change in attitudes. Participants in training scored better on a multi-choice knowledge test on CRM principles than controls.	Trained crews performed better than untrained crews as measure by the Target Acceptable Response to Generated Events or Tasks (TARGETs). 15% more during pre-flight brief, 9% more during high-workload segment.	
	27 US Naval helicopter pilots (12 serving as controls).	Topics covered in training included: Introduction Decision making Assertiveness Mission analysis Communication Coordination Leadership Adaptability Situation awareness	Strong endorsement of training.	No significant difference between trained and controls in attitudes as measured by the ACAQ. Trained aviators scored significantly better on the multi-choice knowledge test than controls.	Comparing trained to control, using TARGETs, trained performed better during pre-flight brief, no difference In low work load times, but engaged in a greater number of teamwork behaviours during high workload segments.	
Stout, Salas & Kraiger (1996), USA.	12 US Navy helicopter pilots (10 serving as a control group) .	Test group received 1 day of training. 5h training organised into 3 modules: Introductory concepts Communication assertiveness	Positive reaction to training.	Positive change in attitudes in 75% of items on CMAQ, but not significant. No significant difference in scores on a multi-choice knowledge test. when compared with controls.	Trained participants performed on average 8% more desired behaviours than control as measured by TARGETs.	

3 Other high reliability industries (n=8 independent studies)

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Brun et al (2001), Norway.	24 cadets from The Royal Norwegian Naval Academy.	The 3 day training covered: Coping in critical situations Communication and decision making Team working It was hoped that the training would increase shared mental models.	Feedback was generally positive	There was no systematic difference between score on a shared mental models questionnaire pre- and post test in the trained group, or when compared to an untrained group	There was a slight tendency for the trained group to out perform the untrained group as judged by experts using a behavioural marker system.	
Byrdorf (1998), Denmark.	Maritime officers from A.P. Moeller and MAERSK Company. Number of subjects not identified.	Training has been delivered for 4 years. Topics covered include: Resource management Assertiveness Communication Team work Stress coping				Over 4 year period of training: Reduction from 1 nautical casualty per 30 ship years to 1 per 90 ship years. Reduction from 1 machinery casualty per 60 ship years to 1 per 90 ship years. 6.5 lost time accidents 1mio. Exposure hour for the fleet of 18 supply vessels to 3.7 lost time accidents 15% reduction in insurance premiums

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
Fonne & Fredriksen (1995), Norway.	432 High speed marine craft navigators from several Norwegian shipping companies.	Two day CRM training course, no details of topics provided	Crews found CRM training to be overwhelmingly acceptable.	Positive attitude change as measured by a 31 item questionnaire. This change was still evident 6 months after the course (although moderated).		
Harrington & Kello (1992), USA	170 US Nuclear Control Room personnel.	Initial CRM training, no details of topics provided.		Positive change in attitudes as measured by the Control Room Operations Attitude Questionnaire (Based on the CMAQ). Increase of 9% for Recognition of stressor effects, 6% for communication and co-ordination and 3% for command responsibility.		
Howard, Gaba, Fish, Yang & Sarnquist (1992), USA	46 US anaesthetists (evaluation feedback from 38).	2 day course. Topics covered included: Dynamic decision making Human performance issues in anaesthesiology Principles of Anaesthesia crisis resource management	The course was highly rated by the participants. Participants rated the course as intense, helpful for their practise of anaesthesiology, and highly enjoyable.	Scores on a written knowledge test showed a significant improvement for residents, but not experienced anaesthetists.		

Author & Country	Participants	Training	Reactions	Learning	Behaviour	Organisation
O'Connor & Flin (under review), UK	77 offshore oil platform workers.	2 day initial CRM training. Topics covered included: Introduction Situation awareness Decision making Communication Team working Personal limitations (stress and fatigue).	The participants were generally positive about the course and its usefulness for working offshore.	Using the Offshore Crew Resource Management Attitude Questionnaire, although not significant, there was a positive shift on decision making, personal limitations No significant effect was found in the ability of participants to recognise the human factors causes of an accident presented as a written scenario.		
Taylor (2000); Taylor (1998); Robertson & Taylor (1995), USA; same data used in all three studies.	Four US airline maintenance companies (responses from 3495 pre-training, 3280 post).	Training not specified in detail, but was not consistent across the companies.	Participants from three of the participating companies showed substantial enthusiasm for the training.	The Maintenance Resource Management / Technical Operations Questionnaire was used to assess attitude change. Positive change in attitudes immediately after training. In one company, follow-up delay survey showed attitudes to participation, teamwork, and stress management did not decay.	Some self reported behavioural change in one company.	In the one company where this was assessed, suggestive evidence for an improvement in both occupational injury and aircraft damage.
Woldring & Isaac (1999), Europe	701 Air traffic controllers from 7 European countries (attitude questionnaire administered to 126 participants)	The 3 day training lasted 3 days and covered: Teamwork Team roles Communication Situation awareness Decision making Stress	Feedback was generally positive as measured using a paper based evaluation form	The 38 item Air Traffic Control Safety questionnaire (based on the FMAQ) showed a significant positive shift in attitudes to the training.		

Section 2

Methods used to Evaluate the Effectivness of Flight Crew CRM Training

Executive Summary

This part of the project was commissioned by the Safety Regulation Group of the UK Civil Aviation Authority (CAA) to identify the methods UK air operators are currently using to assess the effectiveness of their flight deck Crew Resource Management (CRM) training programmes.

Questionnaires and structured interviews were used to obtain information from the industry. A total of 113 questionnaires was returned, a 65% response rate. A representative sample of 20 interviews was carried out with companies who had completed the questionnaire. Kirkpatrick's (1976) hierarchy of training evaluation was used to categorise the types of evaluation techniques that were used.

The primary findings of the study are as follows:

- All of the large companies, and approximately 50% of medium and small companies are carrying out an assessment of flight crew reactions to CRM training. However, the majority of medium and small companies rely on informal feedback.
- Approximately 20% of large and small operators, and 40% of medium operators carry out an evaluation of the attitudes of flight crew to CRM training. All of the large operators who undertook an assessment of attitudes use a questionnaire whilst the majority of medium and small operators rely on oral feedback.
- Approximately 10% of large operators, 20% of medium sized operators, and 40% of small operators carry out an evaluation of the CRM knowledge of flight crew. Generally, this assessment is informal, with about 30% of medium and small operators carrying out a formal knowledge test and the large operators relying solely on oral feedback.
- Approximately 80% of large and medium sized operators and 40% of small operators carry out an assessment of flight crew CRM behaviours. The majority of large operators carry out a formal assessment of behaviour using either behavioural marker systems or technical checklists, with a minority relying on informal feedback. However, the reverse is the case for medium and small operators.
- Approximately 40% of large operators, 50% of medium sized operators, and 20% of small companies reported carrying out an evaluation of CRM training at the organisational level; all are carried out formally.

Conclusions

The survey showed that most UK operators do attempt to evaluate the impact of their CRM training in one way or another. However, many of the methods used would appear not to be based on formal evaluation techniques, and would not provide sufficient information to assess whether CRM training was actually transferring to the flight deck. The main reasons why companies are not evaluating CRM training are i) a shortage of resources (time and personnel) and ii) a lack of guidance on suitable techniques for evaluating training. From these conclusions, it is possible to make a number of recommendations.

Recommendations

- There is a need for guidance on how to evaluate CRM training.
- The evaluation of CRM training should be a continuous process.

Two other recommendations, although not directly relevant to the evaluation of CRM training, are also worthy of consideration:

- Companies could be encouraged to pool their limited resources for CRM training and evaluation.
- The content of CRM training for single pilot crews should be examined.

1 Introduction

1.1 Overview

This part of the project identifies the methods UK air operators are currently using to assess the effectiveness of their Crew Resource Management (CRM) training programmes for flightdeck crew.

1.2 Aim

The aim of the study was to gain an understanding of the techniques that are being used by UK air operators to evaluate the effectiveness of the CRM training given to commercial flight crew.

1.3 Background

The CAA requires that CRM training be undertaken annually by commercial flight crew (CAA, 1998a, b). A recent survey of International Air Transport Association (IATA) affiliated airlines indicated that 96% of respondents were running CRM courses. Over 60% of these courses had been in existence for five years or more (O'Leary, 1999).

In recent years a number of regulatory bodies have become increasingly concerned with the evaluation of the effectiveness of CRM training. The CAA has stated that 'The variability of CRM standards and the lack of common practical reference criteria have indicated the need for research into means of assessment' (CAA, 1998a). The US Federal Aviation Administration (FAA) also recognises the crucial role of CRM evaluation: "It is vital that each program be assessed to determine if it is achieving its goals" (FAA, 1993) and further, "each organisation should have a systematic assessment programme" (FAA, 1998, 14). Therefore, this project examined the literature to scrutinise the techniques that have been used to evaluate the impact of CRM training, and then investigated the methods which UK air operators are actually using to evaluate the effectiveness of their own CRM training.

2 Method

It was decided to use two different techniques (questionnaires and interviews) to obtain the required information from the operators. There are advantages and disadvantages to both methods, and by using both the strengths of each can be consolidated.

2.1 Questionnaire survey

2.1.1 Questionnaire development

The questionnaire was developed on the basis of the literature review and a previous survey conducted at a workshop on human factors training evaluation at the Australian Aviation Psychology Conference, in Manly, Australia in November 2000[1]. Generally the questionnaire distributed at the Manly conference appeared to be well understood, and the information obtained was found to be useful. However, as a result of the comments received by the respondents, a number of changes were made to develop the Manly questionnaire for the current study. The instructions were made more explicit, an explanation of the meaning of each level of Kirkpatrick's hierarchy was given, and it was indicated that the respondents should answer with reference to both basic/foundation training and recurrent CRM training for flight crew. Also, the respondents were given the opportunity to write any comments on each

1. The details of the survey are available on the internet at www.psyc.abdn.ac.uk/homedir/poconnor/AAvPA%20survey.pdf

section of the questionnaire. The revised questionnaire consisted of nine sections as shown in Appendix 3.

The first section of the questionnaire is concerned with whether flight crew received basic/foundation and recurrent CRM training, and who provided the training.

The next five sections of the questionnaire are concerned with whether an evaluation was carried out at any of the four levels of Kirkpatrick's evaluation discussed in Section 1 of this report, i.e. reactions, learning, (split into attitudes and knowledge), behaviour, and organisation. In each of these sections, questions related to the type of CRM evaluation data that were collected by an operator, the method adopted to do this, and how the information was used to evaluate CRM training.

Section 7 is concerned with establishing whether any other evaluation techniques were employed which do not fit into the four Kirkpatrick categories already identified.

The purpose of section 8 is to gather information about why companies are not evaluating CRM training. Respondents are asked to rank the reasons for not evaluating CRM training. (The information obtained from the questionnaire used at the Manly conference was employed to develop the items in this section).

Section 9 asks respondents to rank in order of their effectiveness methods of evaluating CRM training. They are also given the opportunity to describe in their own words what can be done to help operators to evaluate the effectiveness of their CRM training. Again, the information obtained from the Manly questionnaire was used to develop the items for ranking.

2.1.2 **Sample**

The target organisations were every UK AOC and PAOC holder. An AOC is issued by the CAA and is required by aircraft operators flying scheduled or charter public transport flights, including cargo, air-taxi and pleasure-flying work. Police forces are issued with a PAOC licensing them to perform their specialist air operations. To ensure the questionnaire reached the appropriate person in each organisation, it was necessary to identify the individual responsible for CRM training for each AOC and PAOC holder. The CAA provided a preliminary list of 159 UK AOC and 30 PAOC holders. A researcher then endeavoured to contact, by telephone, every company in order to acquire the name of the individual within the company responsible for CRM training (even if the training is contracted out). The contact details for 144 of the AOC holders and all 30 PAOC holders were obtained.

2.1.3 **Method of distribution.**

The questionnaires were sent out by post. The return rate for this method of survey can be very low (10% to 20%) and some researchers have confidence in a postal survey only when the return rate is at least 50% (Heiman, 1995). To ensure that the return rate was as high as possible, the questionnaires were sent to a named individual responsible for CRM training within the particular organisation. For the 15 companies where it was not possible to establish the individual responsible for CRM training, the questionnaire was sent to the Chief Pilot. Each person received a return envelope addressed to Aberdeen University, cover letter (Appendix 1), instructions (Appendix 2) and questionnaire (Appendix 3). All the questionnaires were sent out at the beginning of June 2001. They were then sent out again with a reminder letter to the 123 companies who had not returned the questionnaire by the end of July. The last questionnaire returned was received on 6 September 2001.

2.2 Interview survey.

Information obtained from the questionnaire was augmented with interview data. The rationale for this was that even with well-designed questionnaires it is only possible to obtain data on the questions that have been asked. Moreover, it is not possible to probe the responses to the questions in more detail. So, supplementary interviews were carried out with a sample of 20 respondents from the questionnaire survey. It was decided to use a structured interview technique for two reasons. Firstly, four researchers were conducting the interviews. Therefore, to reduce the likelihood of interviewer bias and ensure that the interviews were as consistent as possible, an interview schedule was used. This was designed using the questionnaire as a basic outline (Appendix 3). Although interviewer bias is potentially always present, using a script for the interview ensures that all of the interviewees are asked the same questions in a consistent, unbiased fashion (Heiman, 1995). Secondly, the structured interview provides meaningful data that are relatively easy to analyse and compare with the responses obtained in other interviews.

A necessary pre-requisite for the companies to be interviewed was that the questionnaire had been completed. The sample was to include interviews with a random selection of companies from each group identified below:

- Major operator
- Charter operator
- Budget operator
- Small operator
- Rotary operator

It was necessary to carry out either face-to-face or telephone interviews with the individuals who had completed the questionnaires. Each interview took approximately an hour. Although face-to-face interviews were preferred, this was not always possible due to the flying schedule of the interviewees. Additionally, copies were requested of the materials that the airlines were using to carry out the evaluation (e.g. behavioural marker checklist).

3 Results

3.1 Flight crew CRM training

Each of the nine sections will be examined in turn with respect to the size of the company. Information gathered from the questionnaire will be presented first and then the information from the interviews summarised.

3.2 Participants

3.2.1 Questionnaire respondents

A total of 113 questionnaires was returned (65% response rate). To make comparisons between the companies, it was decided to differentiate between large, medium, and small companies. Following discussions with the CAA, it was decided to use the capacity in 'Percentage of all available Seat Kilometres' between April 2000 to March 2001 (CAA, 2001) and the type of licence held.

Available Seat Kilometres represents the number of seats available for passengers multiplied by the number of kilometres those seats are flown. The type of licence refers to whether the operator holds a Type A or Type B licence. Type A licence holders are permitted to carry passengers, cargo and mail on aircraft with 20 or more seats, and Type B licence holders are permitted to carry passengers, cargo and mail on aircraft with fewer than 20 seats and/or weighing less than 10 tonnes (CAA, 2001). The method of categorisation of companies into large, medium and small was as follows:

- Large: Type A licence operators with 1% (or greater) of the percentage of all available seat-kilometres.
- Medium: Type A licence operators with less than 1% of the percentage of all available seat-kilometres.
- Small: Type B licence operators and Police Air Operating Certificate holders (PAOC).

The demographic information is summarised in Table 2.

Table 2 Sample demographics (number of respondents in brackets).

Type of certificate AOC holders…. 57% response rate (90) PAOC holders…73% response rate (22) No details …. 1		
Type of licence (% of total sample) Type A…. 30% (34) No details …. 1% (1)	Type B…. 50% (56)	Police …. 19% (22)
Aircraft Type Fixed-wing…. 59% (67)	Rotary…. 40% (45)	No details …. 1% (1)
Size of Company Large…. 10% (11) No details …. 1% (1)	Medium…. 20% (23)	Small …. 69% (78)

3.2.2 Interviews

A total of 20 interviews was carried out by four researchers between July and September 2001 (10 face-to-face interviews, and 10 telephone interviews). Using the same metric for calculating the size of the airline as for the questionnaire survey, nine interviews were carried out with large companies, seven with medium sized companies, and four with small companies. These companies consisted of scheduled, charter, and budget airlines. This sample also included a number of rotary wing operators, including PAOC holders.

3.3 Flight crew CRM training

3.3.1 Flight crew CRM training: Questionnaire findings

For the large operators, basic/foundation and recurrent training are almost exclusively provided by an in-house training department. This figure falls to 74% (17 respondents) for the medium sized companies, with the remainder of them employing specialist consultants to carry out the training. For the small companies, half the basic/ foundation training is carried out by specialist consultants, with the remainder split almost equally between in-house training departments and other airlines employed to carry out the training. However, for the recurrent training, 56% (44) of the small companies carried out this training in-house, with only 9% (7) of them using another airline and 15% (12) employing specialist consultants.

For the large and medium companies 73% (8) and 57% (13) of the CRM training respectively was designed for the specific operations of the flight crew. A number of participants commented that, although the initial CRM training is generic, the recurrent training is designed specifically for the type of operations carried out by the company. For the small companies almost 50% (38) of the training was generic and not designed specifically for the operations carried out by the crews. It was pointed out by a number of the small operators that the CRM training their crews were receiving was not entirely appropriate as it was designed for multi-crew operations in large passenger aircraft. Comments from respondents included:

"CRM training has limited relevance to our crew who fly single crew, light twin-engine aircraft."

"I am not convinced of the value of CRM training in a single pilot environment."

"CRM training is a multi-crew tool in general. We operate single pilot with passengers where CRM has limited application, but we try!"

"More effective programme designed specifically for small light aircraft required."

3.3.2 Flight crew CRM training: Interview findings

The findings from the interviews largely endorsed the information obtained from the questionnaire responses. The reasons given in the interviews for the benefits of carrying out the training in-house were:

- It allows courses to be designed for specific operations and cultures of the airline.
- It is the most cost effective method of delivering the training.
- If there is a particular pattern of problems occurring, then the training can be customised to address the issues.

3.4 Assessment of reactions to CRM training

3.4.1 Assessment of reactions to CRM training: Questionnaire findings

All 11 of the large operators reported that they carried out an assessment of the reactions of the participants to the training. For the medium and small operators, this figure reduced to just over 50% (14 and 44 respondents respectively). There were also differences in the methods used by the different sized operators to assess the reactions of the participants. Almost 75% (8) of the large operators reported that they used a reaction sheet to obtain feedback. However, only 29% (4) of medium sized operators, and 14% (6) of small operators who were carrying out any evaluation at this level used reaction sheets. The companies who were not using a reaction sheet relied on oral feedback from the participants. It was commented that this information was not only obtained after training but whenever the instructors came into contact with the crews.

3.4.2 Assessment of reactions to CRM training: Interview findings

Those companies that used reaction sheets had generally developed the questionnaire in-house and the course participants filled them out anonymously. The reaction sheets were normally fairly short, and consisted of closed questions on each of the topics covered in the course. Participants were given the opportunity to rate the training using a Likert scale (for example ranging from 1, '*very poor*' to 9 '*very good*'). Space was also available in the reaction sheet for the participants to add their own comments. Although the data tended not to be formally analysed to identify trends, the information was examined and used to make changes to the training course if required. In some companies, however, it transpired from interviews that the 'reaction sheet' was a blank piece of paper where the participants were asked to record any comments.

Companies using oral feedback from participants considered this to be sufficient for course instructors to assess how the training was being received and to establish what changes needed to be made to the course. In addition, as commented on in the questionnaire, feedback was obtained not only after the training, but whenever instructors came into contact with the crews at other times. All of the companies interviewed stated that, in general, CRM training was received positively by most of the participants.

3.5 Assessment of attitudes to CRM training

3.5.1 Assessment of attitudes to CRM training: Questionnaire findings

Of all respondents to the questionnaire, 21% (24 respondents) reported carrying out an evaluation at the attitude level. Looking at the three different sizes of company separately, 18% (2) of large companies, 43% (10) of medium companies, and 15% (12) of small companies assessed flight crew attitudes to CRM. One of the large companies used a company specific attitude questionnaire, the other used the Cockpit Management Attitude Questionnaire (CMAQ) designed by Gregorich, Helmreich, and Wilhelm (1990). These measurement techniques were used by 20% (2) of the medium sized operators, and only 8% (1) of the small operators that carried out an assessment of attitudes. Thus, as with the reactions to the training course, the majority of the respondents rely on informal oral feedback to make an assessment of the attitudes of flight crew to the concepts covered in CRM training. In addition, of those companies who carry out an evaluation of attitudes to CRM concepts, only 75% (18) have used this information to evaluate the company's CRM training.

3.5.2 Assessment of attitudes to CRM training: Interview findings

Only two (10%) of the companies interviewed carry out a formal assessment of the attitudes of the aircrew. One of the companies used an adaptation of the CMAQ. The course participants completed the survey anonymously. However, the information has not been used, as yet, to change the training (although there is the intention to do so). The other company had an attitude questionnaire designed specifically for the company and the course participants complete it anonymously.

The response from the companies who did not carry out a formal assessment of attitudes was to question the relevance of this type of analysis, the lack of suitable measurement systems and the fact that it is "another form to have to fill out". The lack of time or expertise needed to carry out an evaluation of attitudes was also identified as an additional complication.

3.6 Assessment of knowledge of CRM concepts

3.6.1 Assessment of knowledge of CRM concepts: Questionnaire findings

The largest proportion of companies who reported carrying out a knowledge assessment were those categorised as small (42%; 33 respondents). The proportion of medium (26%; 6) and large companies (9%; 1) was much smaller (one company who did not report their size also reported performing out a knowledge evaluation). The most common technique used by all the companies carrying out a knowledge test was oral feedback from flight crew (88%; 36), with multiple choice tests and written exams accounting for 12% (5) in each case. A total of 61% (25) of the companies who carried out a knowledge assessment reported using the information obtained from the feedback to evaluate their CRM training.

3.6.2 Assessment of knowledge of CRM concepts: Interview findings

Three of the companies interviewed carried out a formal assessment of flight crews' knowledge of the concepts covered in CRM training. One of the medium sized companies required flight crew to complete a written exam two weeks prior to each annual CRM training period. This test was developed in-house and is open book and non-jeopardy. This approach allows any common weaknesses to be identified and addressed by a change in the CRM training. Two small companies also reported that they used a formal assessment of knowledge. One of the companies tested flight crew every six months. This test was developed in-house and is also open book and non-jeopardy. It has not been running long enough to draw any firm conclusions. The other company uses a self-assessed multi-choice test, also developed in-house. This

information is retained as part of crew records. However, the majority of companies interviewed who carry out either an informal or no assessment of knowledge reported that a formal assessment would be unacceptable to their flight crew.

3.7 Assessment of CRM behaviour

3.7.1 Assessment of CRM behaviour: Questionnaire findings

A total of 82% of large operators (9 respondents) report that they have carried out an assessment of flight crews' CRM skills in the past two years. For the medium sized operators, this was the case for 70% (16) of respondents, and 44% (34) for smaller operators. One company who did not report their size also reported performing an assessment of CRM behaviour.

Behavioural marker systems were used by 67% (6 respondents) of large companies, with 33% (3) relying on informal feedback. For medium sized and small operators who conducted a behavioural assessment, 19% (3) and 12% (4) respectively used behavioural markers, 13% (2) and 24% (8) reported using technical checklists, and 67% (11) and 65% (22) used informal feedback for both sizes of operators. Examining the 14 operators that used behavioural markers, 64% (9) used company specific behavioural marker systems, with 22% (3) using NOTECHS (Avermaete & Kruijsen, 1998; see Appendix 8), and 14% (2) using the Line/LOS checklist (Helmreich, Wilhelm, Kello, Taggart & Butler, 1990; see Appendix 9).

Companies carrying out an evaluation of flight crew behaviours were asked when the assessment took place. The responses for the large and medium companies were similar with a fairly equal split between base/proficiency checks, line/route checks, and simulator/Line Oriented Flight Training (LOFT). The majority of large carriers carried out an evaluation at all three occasions (78%; 7 respondents), whereas only 50% (8) of the medium carriers did so. For the 34 small companies who carried out CRM behaviour assessment there was an almost even split between base/ proficiency checks and line/route checks. The majority (50%; 17) of small companies carried out the evaluation on only one of the occasions listed. The three small and one medium company which selected the 'other' response category provided in the questionnaire reported that the CRM skills of the flight crew were discussed after every flight. It was also reported that 48% (29) of all companies assessing behaviour use the information about the CRM skills of flight crew to evaluate the training.

One operator also identified a behavioural level evaluation technique called a 360° appraisal. A follow-up telephone conversation with the Chief Pilot provided further information on the process which was used. An outside expert developed a questionnaire in which the managers, pilots, and crew were given the opportunity to provide ratings of their own performance, as well as those of the other members of the company (there are only a small number of employees). This allowed for the identification of potential mismatches between self-perceived performance and 'real' performance. The process provided the individuals with the opportunity to receive feedback on their skills, and how their perception of them differed from those of their work colleagues. Moreover, using several raters from different sources increases the reliability and validity of performance evaluations. The resulting feedback is extremely comprehensive, as using multiple raters captures several aspects of the individuals' performance. However, a 360° feedback exercise needs to be carefully managed (British Psychological Society, 2001). It will only be effective in organisations that have a learning or proactive culture that is open to sensitive appraisal and that are willing to learn and improve.

3.7.2 Assessment of CRM behaviour: Interview study

Seven of the companies interviewed reported using a behavioural marker system. Two reported carrying out a formal assessment using systems based on the NOTECHS framework (see Appendix 8 for details of the system). In one of these companies, the results of the assessment were fedback informally to the crews after base/proficiency checks and line checks; and in the other, the results were recorded in the flight crew training records. The training captains undertaking the assessments had no formal training in using the system but they felt able to carry out the ratings accurately.

Three of the seven companies used systems based on the University of Texas Line/ LOS Checklist (UTLLC; Appendix 9). Companies using it appear not to have provided formal training for raters. The information is retained as part of a crew member's training record. This evaluation occurred at base, on the line, and/or during simulator training. The other two companies were using behavioural marker systems developed in-house (by training captains) at line and simulator checks. Their assessments are recorded in the flight crew training records.

Three other companies used technical checklists that included one or two items relating to CRM skills. The remaining companies interviewed carried out an informal assessment of behaviours using oral feedback from the trainers after simulator/LOFT training. The reason given for not using behavioural markers was that the trainers found this type of system complicated and would be unable to use it accurately. Also, it was stated by some respondents that many operators would not carry out this (or any) evaluation unless it was required by the regulator.

3.8 Assessment of organisational effects of CRM training

3.8.1 Assessment of organisational effects: Questionnaire findings

In 36% (4) of the large companies surveyed, 52% (12) of the medium companies and 27% (21) of the small companies an evaluation was carried out at the organisational level. The respondents were given a list of six evaluation methods, and were asked to indicate if they were carrying out any evaluation at this level (Appendix 4). However, a number of the techniques were not specific to the flight crew alone, such as company climate surveys, and business performance measures and these may have limited utility for the evaluation of CRM training. Other measures which were being used, such as safety performance, incident reporting, confidential reporting, CRM training audit, and technical performance, have more direct relevance to the flight crews and CRM training (Table 3).

Table 3 Organisational performance evaluation techniques reported to have been used (in percentages).

	Size of company		
	Large	**Medium**	**Small**
Safety performance	30	23	18
Incident reporting	10	23	25
Confidential reporting	20	13	18
Technical performance	30	7	5
CRM training audit	0	7	9
Non-flight crew specific evaluations	10	27	25

The largest proportion of the 37 companies who collected information at an organisational level reported using one technique only (41%; 15 respondents). The most commonly used techniques were safety performance data, incident and confidential reporting. The information has been used to evaluate CRM training by 41% (15) of the companies who collected organisational data.

3.8.2 Assessment of organisational effects: Interview study

Although none of the small companies interviewed reported carrying out an evaluation of CRM training at an organisational level, four of the large companies, and four of the medium sized companies reported carrying out such evaluations. However, in the interviews, it was found that three of these eight companies were referring mainly to business and internal audits with limited implications for CRM training.

The five companies which reported carrying out an evaluation at this level which was relevant to flight crew collected information such as confidential reporting, incident reporting, safety performance, and technical performance (e.g. information from flight data recorders). This information is used to effect changes in CRM training, particularly if trends are detected. Two companies reported that they used the British Airways Safety Information System (BASIS). One of the companies reported that they had only recently started using it, so sufficient data were not yet available to make an adequate assessment. However, they considered the information would be valuable for identifying issues, and that CRM training may be a medium which could be used to address problems. BASIS is reputed to be a most popular aviation safety management tool and is used by over 150 organisations including major airlines, regulatory authorities and aircraft manufacturers (British Airways, 2001).

3.9 Use of other evaluation techniques

None of the companies surveyed or interviewed reported using any other novel method of evaluation not listed in the questionnaire.

3.10 Reasons why CRM training effectiveness is not evaluated

3.10.1 Reasons why CRM training effectiveness is not evaluated: Questionnaire findings

The responses given to the questions about why CRM training is not evaluated were common across all company sizes. When asked to choose the factors that prevent companies from evaluating CRM training, the most common choices (Appendix 4) were time (26%; 60 respondents) and resources (25%; 58), followed by expertise (18%; 42), availability of measurement systems (19%; 44) and management support (10%; 24). However, a number of respondents also commented that effectiveness was not evaluated as this was not required by the regulator.

Respondents were asked to rank a number of possible problems relating to the evaluation of CRM training. The overall scores for each of the methods were calculated as follows: the number of respondents that ranked a method in position 1 was multiplied by 3, the number who ranked it in position 2 was multiplied by 2, and these numbers were added to the number of respondents who ranked it in position 3. This gave an overall score for each method. To illustrate, 42 respondents ranked availability of measurement systems as the best method (position 1), 22 ranked it as second, and 12 ranked it as third. Therefore, the overall score for base/proficiency checks is (42x3)+(22x2)+12= 182. The overall scores were then ranked in order from 1 to 7 (Table 4). It should be noted that a limitation of this question was that time was

omitted as a possible problem. Nevertheless, the question provided useful insight into some of the difficulties identified by the respondents.

Table 4 Ranking of problems relating to the evaluation of CRM training (1= greatest problem).

	Large	Medium	Small	All
Availability of measurement systems	1	1	1	1
Competence/expertise of evaluators	2#	2	3#	2
Quality of measurement systems	=3	3#	2	3#
Financial	6	5	4	4
Flight crew's attitude/acceptance	=3	4	5	5
Management	5	6	6	6
Other	7	7	7	7

#Indicates a large drop in the overall scores after this point.

Table 4 shows clearly that the main problems for all operators relate to the availability of measurement systems, competence of evaluators, and the quality of measurement systems. There were differences between company size as to whether evaluating CRM training is judged to be useful. For large companies, 82% (9) of respondents considered evaluation to be beneficial, for medium sized companies 70% (16) but only 46% (36) for small companies. The majority of comments favoured CRM evaluation in that it was perceived to be useful in facilitating changes to training and in providing feedback to trainers, etc., for example:

"Evaluation of any training is necessary for change and progress."

"Without evaluation you cannot improve training systems/techniques."

"Evaluation is necessary to support the rationale for training, measure the success of training, and to identify weak performance areas."

However, some respondents thought that it was of limited use to their organisation as, for example, their company was a single pilot operation (see paragraph 3.3.1).

Many respondents (47%, 53) gave suggestions about what could be done to help companies to evaluate CRM training. The most common theme was that guidance was required from the regulator on how the evaluation should be carried out. The following is a representative sample of comments.

"We need more guidance from the regulator in order to enhance any CRM training. Training costs have to be justified, and this can only be done with help from the CAA."

"The CAA needs to provide guidance material which would assist in devising a simple, plain language assessment tool."

"This is a science still in its infancy and any clear guidance would be welcome."

The other main area commented upon was the need for effective assessment instruments which are reliable, affordable, and simple to use. The following are examples of responses:

"Effective measurement systems would be very helpful."

"There is a need for the availability of easy to use measurement systems."

3.10.2 Reasons why CRM training effectiveness is not evaluated: Interview study

The interview responses supported the evidence obtained from the questionnaire study. The greatest perceived problem reported by the companies relating to CRM evaluation was the non-availability of measurement systems. The people carrying out the training are generally line pilots and, whilst interested in the subject of CRM evaluation, they do not profess to be experts in it. Consequently they are concerned that they may not have the ability or time to analyse and use data obtained from an evaluation system. Difficulty in finding and using a suitable evaluation system was reported by a number of companies, which led to a request for more guidance and expert help. It was commented that, if the CAA want to ensure that CRM training is evaluated, this must be prescribed in the regulations and quality tools must be made available to carry out the task.

The smaller companies were less supportive of the benefits gained by evaluating CRM training (only 46%, 36 respondents, thought it was beneficial). There was also scepticism in these companies that it is possible to evaluate CRM training at all. There appears to be a consensus that CRM training is generally a multi-crew training mechanism and of limited relevance to single crew aircraft. Nevertheless, in the large and medium companies the respondents were more positive about the benefits of CRM training (82%, 9 respondents; and 70%, 16 respondents respectively). Reasons given by respondents for the benefits of CRM training evaluation were that it was important to ensure that the training continues to be effective and to ensure that the company is getting a return on investment for the money spent on it.

3.11 Methods to support CRM training evaluation

3.11.1 Means of supporting the evaluation of CRM training: Questionnaire findings

To gain an understanding of the most preferred method for evaluating CRM training, respondents were given the opportunity to rank three methods of evaluation from a list of 13. The same technique for calculating the overall score was used as at 3.10.1. A summary of the overall scores ranked from 1 to 13 is shown in Table 5.

Table 5 Ranking of methods of evaluating CRM training (1= best method).

	Large	Medium	Small	All
Line/route checks	2	1	1	1
Base/proficiency checks	3#	=4	2#	2
Simulator/LOFT checks	1	2#	4	3#
Self/peer/360° appraisal	=10	=6	3	4
Confidential reports	=7	=4	5	5
Feedback questionnaire	9	3	7	=6
Interview sessions	=10	9	6	=6
Attitude surveys	4	=6	10	8
Knowledge assessment	5	10	=8	9
Incident reports	=7	8	=8	10
Accident data	=10	11	11	11
Technical performance	6	13	13	12
Other	13	12	12	13

# Indicates a large drop in the overall scores after this point.

From Table 5 it can be seen that respondents across all companies believe that an evaluation of CRM training should be carried out on the behaviour of the flight crew during either real or simulated flight. Other types of assessments carried out, such as the reaction (feedback questionnaires), learning (attitude surveys, interview sessions, knowledge assessment) or at organisational levels (confidential reports, incident reports, accident data, technical performance) were chosen with a much lower frequency.

3.11.2 Methods to support CRM training evaluation: Interview study

The preferred methods, identified by the large and medium companies, for evaluating CRM training were simulator/LOFT checks, line/route checks and base/proficiency checks. These were chosen because the assessment is carried out in a realistic environment in which the trainers can actually see how the flight crew are behaving. The small companies often do not use simulators and felt that the assessment should be carried out on the flight deck. All those interviewed thought that it was important that the evaluation be directly relevant to the job. Therefore, as with the questionnaire respondents, the interviewees considered that an analysis of behaviours on the flight deck or in the simulator was the most important method of evaluating CRM training.

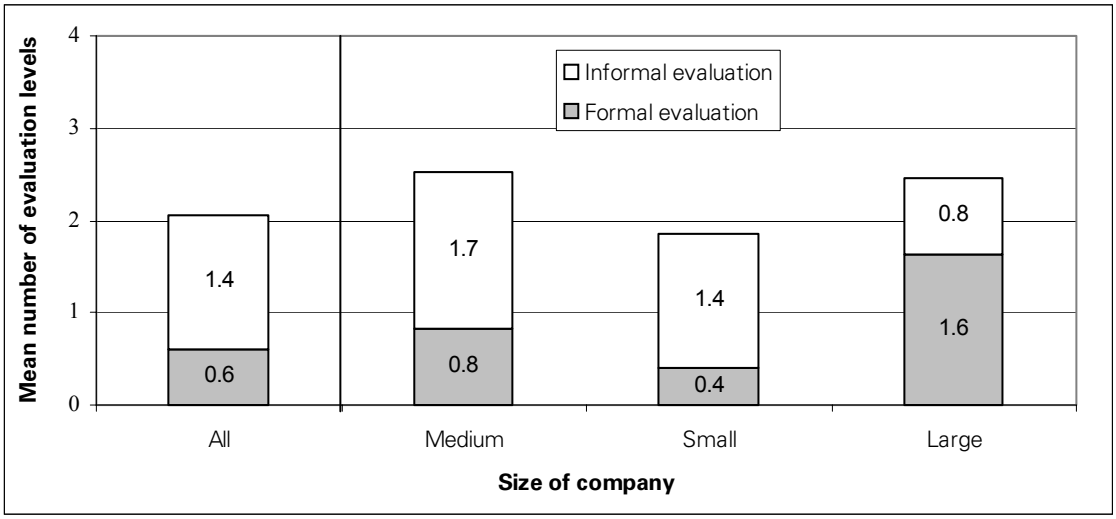
3.12 Summary of Results

The overall conclusion, which can be drawn from the questionnaire survey and interviews, is that the majority of companies are making attempts to evaluate CRM training. Only 14% (16) of companies reported no evaluation of CRM training. All 11 of the large companies, 83% (19) of medium companies, and 83% (65) of small companies carried out some type of CRM evaluation. However, the majority of companies did not perform formal evaluations of CRM training at more than one level of analysis (Table 6 and Figure 2). Further, the proportion of companies carrying out formal evaluations of CRM training decreases for the smaller the company.

Table 6 Number of levels of evaluation carried out by companies (in percentages).

	% of companies performing evaluations		
Number of levels of evaluation	**Formal**	**Informal**	**Formal or Informal**
No evaluation carried out	59	26	14
Evaluation made at only 1 level	29	32	26
Evaluation carried out at 2 levels	5	25	27
Evaluation carried out at 3 levels	7	8	15
Evaluation carried out at 4 levels	0	7	10
Evaluation carried out at 5 levels	0	2	8

Figure 6 Number of evaluation levels carried out separated by company size.



Figures 7 to 10 allow an examination of the proportion of informal and formal evaluations of CRM training reported to be carried out by the three sizes of company. These graphs also illustrate that a greater proportion of large companies carry out formal evaluations of training and that an assessment at the reactions and behaviour levels are the most frequently performed across all three sizes of company.

Figure 7 Percentage of *a*ll companies carrying out evaluations at each level of the hierarchy.

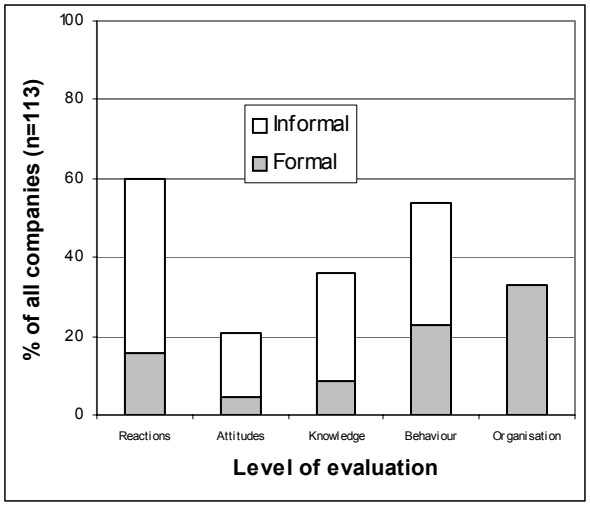


Figure 8 Percentage of large companies carrying out evaluations at each level of the hierarchy.

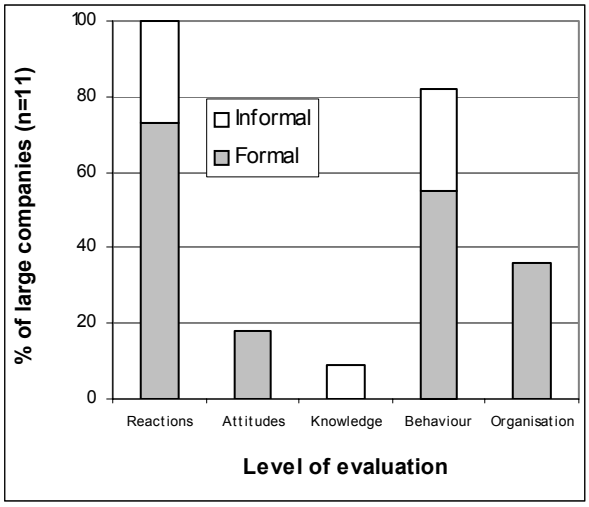


Figure 9 Percentage of medium companies carrying out evaluations at each level of the hierarchy.

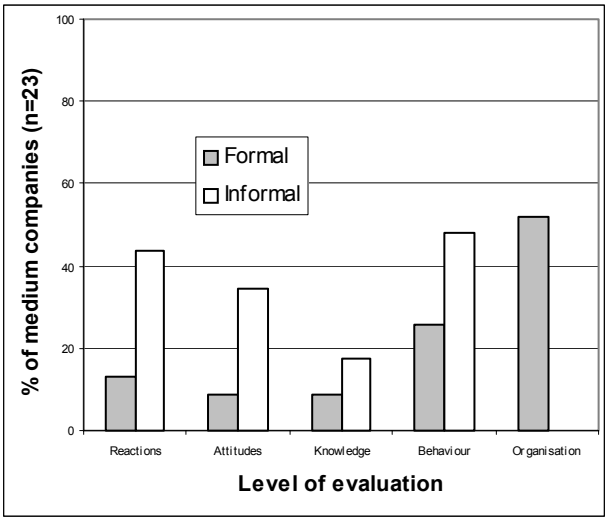
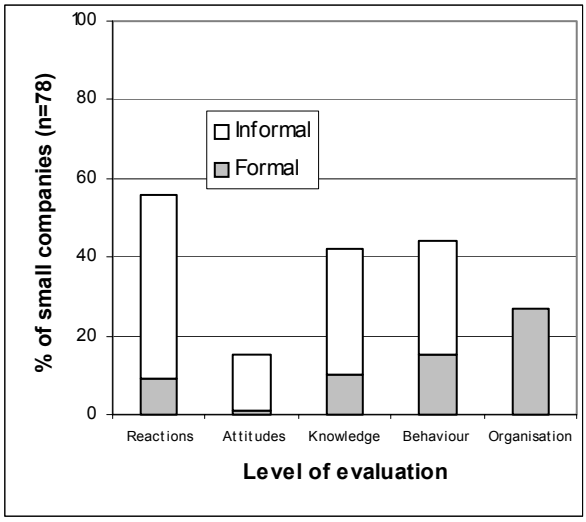


Figure 10 Percentage of small companies carrying out evaluations at each level of the hierarchy.



4 Discussion

This section reviews the results from each of the nine topics identified in the analysis, before making a number of recommendations relating to CRM training evaluation.

4.1 Flight crew CRM training

The results from both the questionnaire and interviews indicate general support for CRM training. However, it is recognised by the industry that for CRM training to be most effective, it should be customised specifically for a particular operator and the type of operations being undertaken. Further evidence of the benefits of this approach are apparent when looking at problems that have occurred where CRM training has not been customised and a particular training package is used in a different country or industry from that for which it was originally developed (Flin & O'Connor, 2001; Wiener, Kanki & Helmreich, 1993). Nevertheless, the survey shows that this is a problem for the small operators and particularly those operating single pilot aircraft.

4.1.1 Single pilot CRM training

The perception of respondents operating single pilot aircraft is that CRM training is primarily relevant for multi-crew operations. However, it is proposed by the authors that CRM is just as critical for the single pilot who must be aware of, and have the ability to manage, every resource available to support safe flight. Although the traditional CRM training topics that emphasise the need for team working on a multi-crew flightdeck are not so relevant to single pilot operations, other aspects of CRM training such as situation awareness, communication, decision-making, and personal limitations are just as crucial. The single pilot simply has a different set of resources and must manage them differently (Harris, 1995). Further, cockpit management for a single pilot may arguably be more demanding than for pilots in a multi-crew cockpit, and it is important that training reflects these differences. Therefore, it is necessary for them to attend CRM training that is specifically designed for the type of operations they carry out. The syllabus must be tailored to their work environment, and any case studies, exercises, etc, must be focused on single pilot operations. Resources on

single pilot CRM training are available on the Royal Aeronautical Society's Human Factors Group web site[1].

4.2 Assessment of reactions to CRM training

An evaluation of the reactions to training is valuable for the trainers as it provides feedback as to its relevance and where improvements can be made. It is also important to establish whether the participants actually liked the course or thought that it was useful, as crew are unlikely to change their attitudes or behaviour if this is not the case (Kirkpatrick, 1998). The survey indicated that feedback is normally obtained either orally or from a reaction sheet.

There are a number of limitations when relying on oral feedback to assess the reactions to training. These are as follows:

- It is not anonymous; therefore people may be unwilling to be too critical.
- The outcome can be hijacked by a few dominant people with points to make which may not reflect the opinions of the silent majority.
- Lack of a structure in the discussion and limited time may result in aspects of the course not being reviewed.

This is not to say that oral feedback should not be used. It should, however, be taken in addition to a reaction questionnaire as it can provide useful information about aspects of the training that have not been covered in the questionnaire. The advantage of using a structured feedback questionnaire over oral feedback is that the former method allows the course participants to make objective comments about the training, possibly, anonymously.

Kirkpatrick (1998: 26) provides the following list of eight guidelines to ensure that trainers get the maximum value from a reaction sheet:

- Determine what you want to find out.
- Design a form that will quantify reactions (include questions that allow participants to rate the course on a numerical scale).
- Encourage written comments and suggestions.
- Obtain 100% immediate response.
- Obtain honest responses (the likelihood of this can be improved by not requiring participants to give their name).
- Develop acceptable standards (establish a level of satisfaction which the training and the instructor must exceed).
- Measure reactions against standards, and take appropriate action (if the training falls below the standard set, act on this to improve the satisfaction level).
- Communicate reactions as appropriate (i.e. ensure that all the personnel involved in training are aware of the reactions to the training).

However, although a positive reaction to the training is desirable, this does not mean, in itself, that the participants have learned anything useful. Nevertheless, a negative reaction almost certainly decreases the chance that any useful learning has occurred.

1. www.raes-hfg.com/

4.3 Assessment of attitudes to CRM training

An evaluation of the attitudes of flight crew to CRM training can be carried out to assess the extent to which the participants have modified their attitudes or values as a result of the training. From the survey and interviews, it was found that the majority of companies did not carry out an evaluation of attitudes and those companies which did relied on informal oral feedback. It is important to measure the attitudes of the flight crew to the concepts covered in the training, as a modification of behaviour is more likely if a change in attitudes has occurred. The limitations of relying on oral feedback of attitudes are the same as for oral feedback on reactions to training (see section 4.2). Indeed, this is particularly the case for attitudes as these can be nebulous concepts which are difficult to verbalise. Therefore, it is considered that companies should adopt a more formalised survey method for assessing attitudes to CRM training. A questionnaire offers an objective and reliable technique for assessing attitude change before and after training and for comparison between different groups. A standard questionnaire that has been developed to do this is the Cockpit Management Attitude Questionnaire (CMAQ).

The CMAQ is a measure of flight crew attitudes "that are empirically or conceptually related to resource management on the flight deck" (Gregorich & Wilhelm, 1993: 180). It was developed solely to assess flight crew attitudes regarding 'interpersonal components' of their job performance and to link these attitudes to behaviour (Gregorich, et al, 1990; Helmreich, 1984). Three factors were derived from 25 Likert scaled items: intra-crew communication and coordination (measured by the Communications and Co-ordination scale); the nature of superior-subordinate roles (measured by the Command Responsibility scale); and recognition of harmful aspects of stress (measured by the Recognition of Stressor Effects scale). It has been used to assess the effectiveness of CRM training, usually at an aggregated level rather than for individual flight crew (see Gregorich & Wilhelm, 1993). However, the CMAQ was designed more than ten years ago when flight crew were receiving first and second generation CRM training (see Helmreich, Merritt & Wilhelm, 1999). Therefore, the CMAQ was tailored to the training in the interpersonal components of the flight crew's job which was covered in these early CRM courses. However, as CRM training has matured, training courses now concentrate on the idea of error being normal and focus on the generation of strategies for its management. Therefore, there is a greater focus on the *cognitive* aspects of the role of the flight crew such as situation awareness, decision-making and workload management, which are not explicitly addressed by the CMAQ.

Therefore, it is important that any attitude instrument is tailored to a particular CRM course so that only those attitudes towards concepts which have been covered in the training are assessed. The items in the CMAQ could be used to aid the formation of a company specific CRM attitude questionnaire as they have been proven to have reasonable psychometric characteristics. However, what is actually required is a number of different sub-scales that have been tested to ensure that they are reliable and discriminatory and which cover all of the possible concepts that are included in CRM training (e.g. team work, leadership, situation awareness, decision-making, communication, and personal limitations). Thus, if a training course consisted of each of these topics, the operator could use the relevant sub-scales to construct an attitude questionnaire specifically for its training course.

4.4 Assessment of knowledge of CRM concepts

Testing students on their recall and understanding of the CRM curriculum can assess learning. This allows an evaluation to be made of the extent to which the participants in the CRM training have acquired and retained knowledge as a result of the training.

The results from the survey and interviews show that if operators carry out knowledge assessment it tends to be on an informal basis. The standard method is to assess acquisition of knowledge using a paper-based test. This could be a reasonably quick and simple way of receiving feedback on knowledge acquisition, and is reported as being used by some operators. However, the questions and answers (both correct and false if using a multi-choice test) must be designed very carefully to avoid either a floor effect (the test is so difficult that the majority of individuals receive very low scores) or a ceiling effect (the test is so easy that the majority of individuals receive very high scores). The questionnaire could be completed by self-scoring, and used as a mechanism for assessing trends over time and identifying where the training required improvement, as opposed to a technique for evaluating the flight crew.

4.5 Assessment of CRM behaviour

Assessment of whether the desired behaviours are being demonstrated is usually carried out by informal observations rather than by using a more formal behavioural rating system. The survey and interview study found that many of the large airlines have developed their own behavioural marker systems and that these are mostly used for training. (See Avermaete & Kruijsen, 1998; Flin & Martin, 2001 for details of some company specific behavioural marker systems). However, the survey also found that the majority of companies (88%) do not use a behavioural marker checklist and if any evaluation of behaviour is carried out it tends to be based on unstructured observations and oral feedback. Whilst this is better than no feedback, there are drawbacks of relying on informal feedback as outlined in the discussion of reactions assessment in section 4.2. Also, a single item relating to CRM training on the technical checklist is not able to provide sufficient detail to allow useful information about performance to be disseminated to the flight crew. Therefore, to ensure detailed feedback is obtained from CRM training, it is suggested that a behavioural marker system is used.

The lack of a widespread use of formal CRM behaviour assessment in Europe has been corroborated by other studies. A survey of 11 major UK airlines in 1997 showed that only five of them had developed a CRM behavioural markers list, and none of these was used for formal CRM assessment (Flin & Martin, 2001). Moreover, in a study including 104 training captains from 14 different European airlines, only 53% were familiar with a behavioural marker system and only 31% had any experience of evaluating CRM skills (O'Connor et al, in press). Many of the smaller companies reported that they do not have the time, resources or expertise to develop their own systems. This indicates a need for a valid and reliable generic behavioural marker system that could be made available to those operators which do not have the resources or expertise to develop their own systems. Such a system would allow instructors to give structured feedback on flight crews' CRM skills, reinforce the importance of such skills and fulfil the need to comply with the recent Joint Aviation Requirement that mandates the assessment of non-technical skills (JAA, 2001). It is anticipated that systems such as the pan-European NOTECHS instrument will help operators to meet these goals. However, the provision of a behavioural marker system alone is not sufficient to assess the CRM skills of flight crews.

What is important is that instructors who conduct an evaluation have been properly trained and calibrated. "*Instructors and check pilots require special training in order to calibrate and standardise their own skills*" (FAA, 1993). This training should be designed to ensure that raters are able to use the behavioural markers accurately and consistently by reducing the likelihood of judgement biases and improving inter-rater reliability (Flin & Martin, 2001). Baker, Mulqueen, and Dismukes (2001) review a number of different approaches to training instructors to assess flight crew non-

technical skills. For example, a research group at George Mason University has produced guidance to assist companies in the development of a CRM skills evaluation system and to train instructors to use it (George Mason University, 1996; Boehm-Davis, Holt & Seamster, 2001). A guidance document on the use of behavioural marker systems has also been produced recently by a group of research psychologists and practitioners from aviation and medicine (Klampfer, et al, 2001).

4.6 Assessment of organisational effects

The objective of a CRM training programme should be to produce tangible evidence of an effect at an organisational level, such as through an improvement in safety and productivity. Therefore, arguably, evidence of an effect at the organisational level is the most valuable evidence of the utility of CRM training. However, it can be very difficult to attribute measured organisational effects only to CRM training given the wide range of other factors which can have an influence (e.g. changes in regulations, organisational restructuring, aircraft type; Gregorich & Wilhelm, 1993).

The Royal Aeronautical Society Human Factors Group (1999) identifies a number of other potential measures of assessing the effects of CRM training. These include: fuel management, punctuality, job satisfaction, insurance costs, and damage to aircraft. Examining trends over time may provide a mechanism for adapting training to address these issues. However, although large companies are able to track such trends, the smaller companies may not have sufficient flight crew or resource for this to be effective. Incident reporting systems, which may be confidential, provide a mechanism for assessing the effectiveness of CRM training and for identifying areas where recurrent CRM training could address such issues. However, adverse events tend to be infrequent and caution should be taken before redesigning a training course on the basis of one or two incidents.

Another measure which could potentially be used to assess the effects of CRM training is information from flight data quick access recorders (FDQAR). These can produce a multi-channel recording of over 100 flight parameters for every second flown (Bannister, 2001). These recordings are played through a special computer programme, such as SESMA (Special Event Search and Master Analysis) used by *British Airways*, which identifies any flight that exceeds any of 56 pre-set tolerances, facilitating further investigation. Thus, this information can be used to assess the frequency with which particular events, such as heavy landings, are occurring in a given company or fleet of aircraft. However, it is important to indicate that it is not possible to assess the human factors causes of the situation (e.g. low situation awareness, poor communication etc.) which can only be inferred. Henderson (2000) suggests that information obtained from simulator and line assessments of pilots should be integrated with information obtained from FDQAR to obtain a greater understanding of the effects of the training and identify areas of training needs.

4.7 Methods to support CRM training evaluation

Both the questionnaire responses and the information obtained from the interviews illustrate that the vast majority of the participants in the survey thought that the flight deck during routine operations is the most appropriate place for carrying out an assessment of the effectiveness of CRM training. However, as described above, the use of CRM rating scales (e.g. behavioural marker systems) is not widespread. Of the 55% of operators that carry out an evaluation at the behavioural level, the majority are relying on either informal feedback or technical checks to assess the CRM behaviour of the flight crew. Thus, while there is already a clear culture within aviation for assessing behaviour on the flightdeck, operators need to be provided with more information on behavioural marker systems (such as NOTECHS or LLC) in an easily usable format. Moreover, specific training is required on the use of the behavioural

markers to ensure that any rating system is being used properly. Resources may limit the assessment of the behaviours of flight crew to the simulator as opposed to actually 'on the job'. However, carrying out a behavioural assessment during line operations will provide stronger evidence that crews are actually using the information learnt during CRM training to behave appropriately and hence effectively.

4.8 Reasons why CRM training effectiveness is not evaluated

The main difficulties for aviation operators in carrying out evaluations of CRM training are that the training personnel have limited time, resource and expertise. The individuals involved in CRM training are generally enthusiastic line pilots who have a particular interest in CRM training. Nevertheless, they often feel that they do not have the skill to carry out an in-depth assessment of CRM training. In addition, there is a lack of guidance on how to carry out evaluation of training and tools to facilitate an evaluation. A useful aid would be a resource pack providing information on how to carry out formal evaluations of reactions, attitudes, behavioural and organisational outcomes. The following points regarding the evaluation of the effectiveness of CRM training arose from the questionnaires and interviews:

- At the reaction level, a sample/master reaction sheet could be designed and made available to be customised to a particular course or training.
- There is not a comprehensive system available for measuring the CRM attitudes of flight crew. As described at 4.3, the CMAQ focuses on flight crew attitudes regarding interpersonal and personal limitations. Consequently, it would be useful to develop several sets of attitude questions relating to the wider range of topics covered in modern, fifth generation (Helmreich et al, 1999) CRM training courses. This would allow the trainers to select the appropriate scales for the topics covered in their training.
- No standard knowledge assessment tool is available. If a measure is to be used, it is important that it is specifically related to the topics covered in the training course. Therefore, there is a need for guidance to instructors on how to design a test for assessing the knowledge of flight crew.
- Behavioural marker systems provide an acceptable technique for carrying out an evaluation of the CRM skills of flight crew (Salas et al, in press). However, with the exception of NOTECHS and the UT LLC (see Klampfer et al, 2001), there is very little information on alternative systems (especially when they have been developed in-house).
- Organisational level assessments, which are directly linked to flight crew performance, are difficult to carry out. Whilst some operators collect information on safety performance and have confidential incident and reporting systems which are used to evaluate CRM training, they are in the minority. Appropriate techniques for carrying out meaningful evaluations at this level still need to be assessed.

To aid the trainers in analysing data obtained from an evaluation of CRM training, there is a need to make this as simple and easy to carry out as possible. The use of spreadsheets for this type of analysis was employed by George Mason University (1996) and Henderson (2000). Spreadsheets can be designed so that the trainer is only required to type in the raw data and the software will carry out the calculations. Thus, using spreadsheets provides a cheap and simple to use analysis system which utilises software packages that most companies will already have installed.

5 Conclusions

CRM training evaluation is important because it identifies whether the time and money spent on the training is having a beneficial effect. It also enables the identification of topics for recurrent training and shows whether CRM training is continuing to improve flight crew performance despite changes in aircraft and crew demographics. The survey findings can be summarised as follows:

- Reactions: All of the large operators surveyed and approximately 50% of medium and small companies are carrying out an assessment of the reactions to CRM training. However, the majority of medium and small companies rely on informal feedback.
- Attitudes: Approximately 20% of large and small operators and 40% of medium operators report carrying out an evaluation of the attitudes of flight crew. Whilst all of the large operators who undertook an assessment of attitudes use a questionnaire, the majority of medium and small operators rely on oral feedback.
- Knowledge: Approximately 10% of large operators, 20% of medium sized operators and 40% of small operators surveyed carry out an evaluation of the knowledge of flight crew. Generally, this assessment was not formal, with only about 30% of medium and small operators carrying out a formal knowledge test, whilst the large operators rely solely on oral feedback.
- Behaviour: Approximately 80% of large and medium sized operators and 40% of small operators carry out an assessment of the behaviours of flight crew. The majority of large operators carry out a formal assessment of behaviour using either behavioural marker systems or technical checklists, with only a minority relying on informal feedback. However, the reverse is true for medium and small operators.
- Organisation: Approximately 40% of large operators, 50% of medium sized operators, and 20% of small operators report carrying out an evaluation at the organisational level. All evaluations are carried out formally, but these were generally not linked back to CRM training.

In conclusion, this survey has demonstrated that, despite the recognition by most respondents that CRM training evaluation is beneficial, few companies are actually carrying out a multi-level, formal evaluation of CRM training. However, it is hoped that more guidance, and better availability of techniques for evaluating CRM training, will aid companies in the design and implementation of the next generation of CRM training.

6 Recommendations

It is possible to make a number of recommendations relating to the evaluation of CRM training.

- There is a need for more guidance on how to evaluate CRM training.

Information regarding possible methods for evaluating CRM training is not readily available. Thus, there is a need for guidance to be made available on methods of evaluation and this should be easy to access and read.

- The evaluation of CRM training should be a continuous process.

The evaluation of CRM training should not occur on a single occasion only. Rather, the evaluation should be carried out continuously to ensure that the training has the desired effect and remains relevant to the operator.

Two other recommendations, although not directly relevant to the evaluation of CRM training, are also worthy of consideration:

- Companies could be encouraged to pool their limited resources for CRM training and evaluation.

Although there is some evidence of companies sharing their resources to aid the development of CRM training, it appears to be infrequent. It is suggested that companies that are carrying out similar types of operations, or use similar types of aircraft, could pool their resources to aid in both the development and evaluation of CRM training. This would be mutually beneficial by providing a cost effective way for companies to maximise the resources they have available.

- The content of CRM training for single pilot crews should be examined.

The survey drew attention to the fact that CRM training methods are generally concerned with multi-crew flightdecks, and seem to have much less relevance to the single-pilot environment. Thus, rather than subjecting single pilot crews to CRM training which they feel is of little relevance, it is crucial that training is designed specifically for these pilots.

Appendix 1 Cover letter

Dear,

CRM Training Evaluation Survey

The Industrial Psychology Group of Aberdeen University has been commissioned by the Safety Regulation Group of the UK Civil Aviation Authority (CAA) to undertake a survey of the methods being used by UK air operators to evaluate the results of their pilots' Crew Resource Management (CRM) training. Our interest at this stage is simply to chart which methods, if any, are being used by air operators to assess the impact of CRM training programmes on pilots' CRM skills and consequent organisational performance. We would be very grateful if you would complete the enclosed questionnaire (it should only take 10-15 minutes) and return it in the envelope provided. If you are not currently conducting any evaluation of your pilots' CRM training, we would still appreciate it if you could complete the relevant items on the questionnaire and return it. If you are not the person responsible for CRM training in your organisation, then please pass the questionnaire to the relevant individual.

Any information obtained for the survey will be de-identified prior to being made available to CAA and individual companies or persons will not be identifiable in any report. The questionnaire responses will be summarised and used to produce a survey report which will be submitted to CAA by the end of 2001. It is anticipated that the CAA will make this publically available as a CAA Paper.

I would be most grateful if you could return your questionnaire by **Friday 29th June** [date change to **Friday 3rd August** in reminder letter] at the latest. If you have any questions regarding this survey then please contact me at the above address or contact Mr Stephen Griffin at CAA SRG (email steve.griffin@srg.caa.co.uk). Further details of the project are available on the Industrial Psychology Group's website www.psyc.abdn.ac.uk/serv02.htm

Yours sincerely,

Rhona Flin

Professor of Applied Psychology

Appendix 2 Questionnaire instructions

CAA Project code: 121/SRG/R&AD/1

Aberdeen University Industrial Psychology Group

Instructions

We would be very grateful if the individual responsible for Crew Resource Management (CRM) training in your organisation could complete this questionnaire. We are concerned with how flight crew CRM training (both basic/foundation and recurrent training) is evaluated in terms of:

- Reactions: an assessment of flight crews' feelings towards the CRM training course.
- Attitudes: modification in attitudes or values after the training.
- Knowledge: the extent to which the participants have acquired knowledge after the training.
- Behaviour: whether the desired behaviour has been achieved as a result of the training.
- Organisation: any organisational change that has occurred as a result of the training.

We are also interested in why you think companies may not be evaluating CRM training, and any additional comments you may have about CRM evaluation. Please return the questionnaire in the envelope provided by **3rd August 2001**.

No individual or company responses will be made available to the CAA and individual companies will not be identifiable in the final report.

Thank you for your help.

For further information, look on the web at: www.psyc.abdn.ac.uk/crmeval.htm or contact:

Paul O'Connor, Industrial Psychology Group, Aberdeen University, Aberdeen, AB24 2UB. **Phone: 01224 273212, fax: 01224 273211, email: p.oconnor@abdn.ac.uk.**

Appendix 3 CAA Interview Schedule

Survey of the methods used to Evaluate the Effectiveness of Flight Crew CRM Training in the Aviation Industry.

AA Project code: 121/SRG/R&AD/1

Aberdeen University Industrial Psychology Group

Instructions for interviewer

1. The interview should take about an hour.
2. When arranging the interview, make sure the questionnaire has been completed by the interviewee, and has been returned to Aberdeen University.
3. Prior to the interview prepare the template interview schedule so that the answers given are marked.
4. The questions to be asked are in bold, the questionnaire questions are in italics, and the questionnaire responses are in normal type face. Questions which are not crucial are marked with a '*' and can be dropped if running short of time.
5. Add any additional questions which may be required to take account of any comments made or to gather information about any unusual evaluation technique, and delete any irrelevant questions.
6. Take a photocopy of their original questionnaire with you to any face-to-face interviews.
7. At the interview thank the interviewee for agreeing to participate. Tell them that we are working on behalf of the CAA, however no individual or company responses will be made available to the CAA and individual companies will not be identifiable in the final report. Remind them that the purpose of the questionnaire was to gain information about the evaluation of CRM training, and the interview is going to go into the responses they gave in the questionnaire in a little more detail. Inform them that the report will be submitted to CAA by the end of 2001, and it is anticipated that the CAA will make this publicly available as a CAA Paper. It is hoped that this project will help to provide CRM instructors with recommendations for methods of CRM assessment to help provide the justification for the costs associated with carrying out the training.
8. While interviewing attempt to get copies of any questionnaires etc. which are used for evaluation.
9. Type up your notes into the template and send it back to Aberdeen University.

Respondent details Name:	
Organisation:	Job title:
Email address:	Telephone number:
Date of interview	Interviewer:

Flight Crew CRM Training

Do flight crew employed by your organisation receive basic/foundation CRM training?

Yes No

2. Do flight crew employed by your organisation receive recurrent CRM training? Yes No

3. Who provides the basic/foundation training? In-house Another airline Consultancy
Why?

4. Who provides the recurrent training? In-house Another airline Consultancy
Why?

6. Is the CRM training course designed for: specific operations Generic
If customised, why?, how?, and by whom?

Comments
Discuss any comments made.

Reactions

7. Is an assessment made of the reactions of flight crew to the CRM training? Yes No
If no why not? THEN GO TO NEXT SECTION. If yes, how often?

8. How is this done? Reaction sheet Oral feedback/debriefing
Can we get a copy?

How was the method developed?

Who completes the information? (i.e. tutor/ participants/both)

Is the information anonymous?

How is the information analysed? (aggregated to allow comparisons to be made across years/groups/instructor)

What is done with the data? (Make changes to the course, evaluate the instructors performance)

What has been found?

Comments
Discuss any comments made

Attitudes

9. Has your organisation carried out an assessment of the attitudes of flight crew to the concepts covered in CRM training within the past two years? Yes No
If no why not? THEN GO TO NEXT SECTION. If yes, how often?

10. How was this done? Company specific attitude questionnaire CMAQ/FMAQ
Can we get a copy?

How was the method developed?

Who completes the information?

Is the information anonymous?

How is the information analysed? (aggregated to allow comparisons to be made across years/groups)

What is done with the data? (evaluate the instructors performance, judge overall satisfaction)

11. Has this information on the attitudes of flight crew been used to evaluate CRM training? Yes No
If yes, what were the findings? If no, have you thought of this?

Comments
Discuss any comments made.

Knowledge

12. Has your airline carried out an assessment of the extent of flight crews' knowledge of the concepts covered in CRM training in the past two years? Yes No
If no why not? THEN GO TO NEXT SECTION. If yes, how often?

13. How was this done? Multiple choice test Written exam Oral feedback
Can we get a copy?

How was the method developed?

Who developed it?

Who completes the information?

Is the information anonymous?

How is the information analysed? (aggregated to allow comparisons to be made across years/groups)

What is done with the data?

(Make changes to the course)

Is the data aggregated to allow comparisons to be made across years/groups?

14. Has the information on the CRM knowledge of flight crew been used to evaluate CRM training? Yes No
If yes, what were the findings? If no, have you thought of this?

Comments
Discuss any comments made.

Behaviour

15. Has your airline carried out an assessment of flight crews' CRM skills within the past two years? Yes No
If no why not? THEN GO TO NEXT SECTION. If yes, how often?

16. How was this done? Behavioural marker system Technical checklist Informal feedback

17. If you are using behavioural marker system, what do you use?

Company specific Line/LOS checklist NOTECHS

Can we get a copy?

How was the method developed?

Who developed the method?

Who completes the information?

Is the information anonymous?

How is the information analysed?
(aggregated to allow comparisons to be made across years/groups, or not actually recorded)

What is done with the data?
(Make changes to the course, evaluate the flight crew, identify performance problems)

18. When does the evaluation take place?
Base/Proficiency checks Line/Route checks Simulator/LOFT

19. *Has the information on the CRM skills of flight crew been used to evaluate CRM training?* Yes No
If yes, what were the findings? If no, have you thought of this?

Comments
Discuss any comments made.

Organisation

20. Have any organisational performance evaluations been carried out within the last two years? Yes No

If no why not? THEN GO TO NEXT SECTION. If yes, how often?

21. What methods of evaluation were used?

22. Has the information collected on the organisation been used to evaluate CRM training? Yes No

Methods used

Company climate survey	Business performance
Safety performance	Confidential reporting
Incident reporting	Technical performance

If yes, what were the findings? If no, have you thought of this?

Can we have a copy?

For each method used, how was the method developed?

Who completes the information?

Is the information anonymous?

How is the information analysed?

(aggregated to allow comparisons to be made across years/groups)

What is done with the data?
(Identify training needs, develop new procedures, assess management initiatives)

Comments
Discuss any comments made.

Other evaluation methods

23. Does your company use any other method to evaluate the effectiveness of CRM training, not described above? Yes No

If no, GO TO NEXT SECTION, If yes,

Can we get a copy?

Why was the method developed?

How was the method developed?

What information does it give the airline?

Who completes the information?

Is the information anonymous?

How is the information analysed?

What is done with the data? (Make changes to the course, judge overall satisfaction)

Reasons why CRM training effectiveness is not evaluated

24. What factors do you think prevent companies from evaluating CRM training effectiveness?

Chosen factors

Time	Expertise
Resources	Availability of measurement systems
Management support	

***Why did you choose these particular factors?**

Factors not chosen

Time	Expertise
Resources	Availability of measurement systems
Management support	

***Why did you choose these factors?**

25. Do you think being able to evaluate CRM training effectiveness would be beneficial and if so why?

Yes No

Why?

Comment
Discuss any comments made.

Methods to support CRM training evaluation effectiveness

26. What do you think are the preferred methods for evaluating whether CRM training has transferred to the flight deck? (1= best method)

Chosen methods

Base/proficiency checks	Line/route checks
Simulator/LOFT checks	Incident reports
Accident data	Confidential report
Technical performance	Attitude surveys
Interviews sessions	Self/peer/360º assessment
Feedback questionnaire	Knowledge assessment

Why did you choose these methods?

How do you think the data from these assessments would help airlines?

Methods not chosen

Base/proficiency checks	Line/route checks
Simulator/LOFT checks	Incident reports
Accident data	Confidential report
Technical performance	Attitude surveys
Interviews sessions	Self/peer/360º assessment
Feedback questionnaire	Knowledge assessment

***Why did you not choose these methods?**

27. What in your opinion, are the main problems relating to the evaluation of CRM training effectiveness?

Problems chosen

Availability of measurement systems Competence/expertise of evaluators Management	Quality of measurement systems Flight crews' attitudes/acceptance Financial

Why did you choose these problems?

Problems not chosen

Availability of measurement systems Competence/expertise of evaluators Management	Quality of measurement systems Flight crews' attitudes/acceptance Financial

***Why did you not choose these problems?**

28. What could be done to help aviation companies to evaluate the effectiveness of their CRM programme?
Discuss any comments made.

Why/why do you not think it is important to evaluate the effectiveness of CRM training?

Do you have any other general thoughts or comments about CRM training evaluation in your company or the industry at large?

General comments
Discuss any comments made.

Appendix 4 Questionnaire responses: all respondents

Background

Response
AOC holders… 90 (55% response) PAOC holders…22 (73% response)
No details …. 1 Total response rate…. 113 (65% response)

Type of licence
Type A…. 30% Type B…. 50% Police …. 19% No details …. 1%

Aircraft
Fixed-wing…. 59% Rotary…. 40% No details …. 1%

Size
Large…. 10% Medium…. 20% Small …. 69% No details …. 1%

Flight Crew CRM Training

1. Do flight crew employed by your organisation receive basic/foundation CRM training?
Yes…. 97% No…. 2% No response …. 1%

2. Do flight crew employed by your organisation receive recurrent CRM training?
Yes…. 86% No…. 14%

3. Who provides the basic/foundation training?
In-house training department. 40.5% Another airline…. 15% Specialist consultancy…. 39%
Other…. 2.5% No training provided…. 2% No response…. 1%

4. Who provides the recurrent training?
In-house training department…. 66% Another airline…. 6%
Specialist consultancy…. 13% Other …. 1% No training provided…. 14%

5. If the training is not provided in-house, please give the name and contact details of the course provider.
Information held by researchers

6. Is the CRM training course designed for:
The specific operations carried out by your flight crew?…. 47% Generic ? …. 39%
Both…. 12% Not applicable…. 1% No response…. 1%

Please add any further comments relevant to CRM training?
Yes…. 37% No…. 63%

1. Use tailored training…. 8%
2. Initial is generic, ongoing is specific CRM training…. 5%
3. Relevance of CRM is questioned (single crew or small organisation) ….5%
4. Six other categories of comments <5%…. 19%

Reactions

7. Is an assessment made of the *reactions* of flight crew to the CRM training?
Yes…. 60% No…. 36% Don't know….. 4%

8. How is this done? (Tick all relevant; applicable responses, n=68)
Reaction sheet…. 26% Oral feedback/debriefing…. 74% Other…. 0%

Comments specific to this section?
Yes…. 20% No…. 80%

1. Informal feedback is provided…. 4%
2. Information is obtained after the training…. 3%
3. Information is obtained at line/base checks…. 3%
4. Information is obtained simulator/recurrent training…. 3%
5. Outcome is implemented in training…. 3%
6. Three other categories of comments <3%…. 4%

Attitudes

9. Has your organisation carried out an assessment of the *attitudes* of flight crew to the concepts covered in CRM training within the past two years?
Yes…. 21% No…. 78% No response…. 1%

10. How was this done? (Tick all relevant; applicable responses, n=24)
Company specific attitude questionnaire….13%
Cockpit/ Flight Management Attitude Questionnaire (*CMAQ/FMAQ*)…. 8%
Other (informal oral feedback)…. 79%

11. Has this information on the *attitudes* of flight crew been used to evaluate CRM training (Applicable responses, n=27)?
Yes…. 75% No…. 25%

Comments specific to this section?
Yes…. 14% No…. 86%

1. Informal feedback provided… 3%
2. The outcome is used to develop the training….2%
3. Does not apply to the organisation / is not perceived as relevant…. 3%
4. Six other categories of comments <3%…. 6%

Knowledge

12. Has your airline carried out an assessment of the extent of flight crews' *knowledge* of the concepts covered in CRM training in the past two years? (If no/don't know, go to question 15)
Yes…. 36% No…. 63% No response…. 1%

13. How was this done (Tick all relevant; applicable responses- first method, n= 41; second method, n= 5)?
1st Method
Multiple choice test…. 12% Written exam…. 12% Oral feedback…. 76%

2nd Method
Oral feedback…. 100%

14. Has the information on the CRM *knowledge* of flight crew been used to evaluate CRM training (Applicable responses, n=41)?
Yes…. 61% No…. 32% Don't know…. 2% No response…. 5%

Comments specific to this section?
Yes…. 13% No…. 87%
1. Does not apply to the organisation / is not perceived as relevant…. 4%
2. Seven other categories of comments <3%…. 9%

Behaviour

15 . Has your airline carried out an assessment of flight crews' CRM *skills* within the past two years? (If no/don't know, go to question 20)
Yes…. 53% No…. 46% Don't know…. 1%

16. How was this done (Applicable responses, n= 60)?
Behavioural marker system ….23% Technical checklist…. 18.5%
Informal feedback …. 58.5%

17. If you are using behavioural marker system, what do you use (Tick all relevant; applicable response, n=14)?
Company specific ….64.3% *Line/LOS checklist*…. 14.3% *NOTECHS* …. 21.4%

18. When does the evaluation take place (tick all relevant; applicable response, n=119)?
Base/Proficiency checks…. 36% Line/Route checks…. 40%
Simulator/Line Oriented Flight Training….20% Other…. 4%

Number of times evaluation is carried on the occasions described above (applicable response, n= 60)
1…. 37% 2…. 31% 3…. 32%

19. Has the information on the CRM *skills* of flight crew been used to evaluate CRM training (Applicable responses, n=60)?
Yes…. 48% No…. 49% No response…. 3%

Comments specific to this section?
Yes…. 14% No…. 86%

1. Provide brief background regarding specifics about logistics of training…. 4%
2. Feedback is provided at an individual level (e.g. line/base/simulator checks)…. 3%
3. It is planned to implement this type of evaluation…. 3%
4. Three other categories of comments <3%…. 4%

Organisation

20. Have any organisational performance evaluations been carried out within the last two years? (If no/don't know, go to question 23)
Yes…. 32.5% No…. 56% Don't know…. 9.5% No response…. 2%

21. What methods of evaluation were used (Tick all relevant; applicable responses, n=83)?
Company climate survey……. 10% Business performance……. 13%
Safety performance…….. 22% Confidential reporting……17%
Incident reporting…… 23% Technical performance…..... 8%
Other (all training audits) …… 7%

Number of different evaluations used (Applicable responses, n=37)
1…. 41% 2…. 24% 3…. 16% 4…. 11% 5…. 3% 6…. 5%

22. Has the information collected on the organisation been used to evaluate CRM training (Applicable responses, n=37)?
Yes…. 41% No…. 59%

Comments specific to this section?
Yes…. 5% No…. 95%

1. Does not apply to organisation / not perceived as relevant…. 2%
2. Three other categories of comments <2%…. 3%

Other evaluation methods

23. Does your company use any other method to evaluate the effectiveness of CRM training, not described above? (If no/don't know, go to question 24, applicable responses, n=1)
Yes…. 1% No…. 99%
Please describe the method.

1. 360^{o} appraisal…. 1%

Reasons why CRM training effectiveness is not evaluated

24. What factors do you think prevent companies from evaluating CRM training effectiveness? (Tick all relevant; applicable responses, n=234)
Time….25.5% Resources…. 25% Management support…. 10% Expertise…. 18%
Availability of measurement systems…. 19% Other…. 2.5%

Number of different factors chosen (Applicable responses, n= 113)
0 …. 15% 1…. 22% 2…. 28% 3…. 17% 4…. 14% 5…. 4%

25. Do you think being able to evaluate CRM training effectiveness would be beneficial and if so why?
Yes…. 54% No…. 15% Don't know…. 12% No response…. 19%

Comments specific to this section?
Yes…. 57% No…. 43%

1. Evaluation is useful / is perceived as relevant…. 30%
2. Single pilot (or single crew) operation…. 6%
3. Does not apply (to organisation) / not perceived as relevant…. 4%
4. Five other categories of comments <4%…. 17%

Methods to support CRM training evaluation effectiveness

26. What do you think are the preferred methods for evaluating whether CRM training has transferred to the flight deck? (Put the top **three** in order, with 1= best method)

Base/proficiency checks	**135**	Line/route checks	**182**
Simulator/LOFT checks	**99**	Incident reports	22
Accident data	9	Confidential report	41
Technical performance	5	Attitude surveys	25
Interviews sessions	27	Self/peer/360º assessment	44
Feedback questionnaire	27	Knowledge assessmen	23
Other	4	No response	44

27. What in your opinion, are the main problems relating to the evaluation of CRM training effectiveness (Put the top **three** in order with 1= greatest problem)?

Availability of measurement systems	**182**	Quality of measurement systems.	**111**
Competence/expertise of evaluators	**112**	Flight crews' attitudes/acceptance	68
Management	36	Financial	71
Other (Time)	6	No response	92

28. What could be done to help aviation companies to evaluate the effectiveness of their CRM programme?

Yes…. 47% No…. 53%

1. External input / guidelines /regulation…. 16%
2. Effective assessment instruments (affordable; reliable; independent)…. 7%
3. 15 other categories of comments <5%…. 24%

General comments

Yes…. 25% No…. 75%

1. CRM is difficult for single crew operations / address specific training for single crew
2. 6%There are problems with CRM or recurrent CRM training…. 5%
3. Six other categories of comments <5%…. 14%

Appendix 5 Questionnaire responses: large carriers

Background

Response
AOC holders… 11 (92% response)

Type of licence
Type A…. 100%

Aircraft
Fixed-wing…. 100% Rotary…. 0%

Flight Crew CRM Training

1. Do flight crew employed by your organisation receive basic/foundation CRM training?
Yes…. 100% No…. 0%

2. Do flight crew employed by your organisation receive recurrent CRM training?
Yes…. 100% No…. 0%

3. Who provides the basic/foundation training?
In-house training department…. 100% Another airline…. 0%
Specialist consultancy…. 0% Other…. 0%

4. Who provides the recurrent training?
In-house training department…. 100% Another airline…. 0%
Specialist consultancy…. 0% Other …. 0%

5. If the training is not provided in-house, please give the name and contact details of the course provider. *Not relevant*

6. Is the CRM training course designed for:
The specific operations carried out by your flight crew?…. 73% Generic ? …. 0%
Both…. 27% Not applicable…. 0%

Please add any further comments relevant to CRM training?
Yes…. 27% No…. 73%

Reactions

7. Is an assessment made of the *reactions* of flight crew to the CRM training?
Yes…. 100% No…. 0%

8. How is this done? (Tick all relevant; applicable responses, n=11)
Reaction sheet…. 73% Oral feedback/debriefing…. 27% Other…. 0%

Comments specific to this section?
Yes…. 9% No…. 91%

Attitudes

9. Has your organisation carried out an assessment of the *attitudes* of flight crew to the concepts covered in CRM training within the past two years?
Yes…. 18% No…. 82%

10. How was this done? (Tick all relevant; applicable responses, n=2)
Company specific attitude questionnaire….50%
Cockpit/ Flight Management Attitude Questionnaire (*CMAQ/FMAQ*)…. 50%
Other (informal oral feedback)…. 0%

11. Has this information on the *attitudes* of flight crew been used to evaluate CRM training (applicable responses, n=2)?
Yes…. 100% No…. 0% Don't know…. 0%

Comments specific to this section?
Yes…. 27% No…. 73%

Knowledge

12. Has your airline carried out an assessment of the extent of flight crews' *knowledge* of the concepts covered in CRM training in the past two years? (If no/don't know, go to question 15)
Yes…. 9% No…. 91%

13. How was this done (Tick all relevant; applicable responses, n= 1)?

Multiple choice test…. 0% Written exam…. 0% Oral feedback…. 100%
Other…. 0%

14. Has the information on the CRM *knowledge* of flight crew been used to evaluate CRM training (Applicable responses, n=1)?
Yes…. 100% No…. 0%

Comments specific to this section?
Yes…. 9% No…. 91%

Behaviour

15. Has your airline carried out an assessment of flight crews' CRM *skills* within the past two years? (If no/don't know, go to question 20)
Yes…. 82% No…. 18%

16. How was this done (Applicable responses, n= 9)?
Behavioural marker system ….67% Technical checklist…. 0% Informal feedback …. 33%

17. If you are using behavioural marker system, what do you use (Tick all relevant; applicable response, n=6)?
Company specific ….100% Line/LOS checklist…. 0% NOTECHS …. 0%

18. When does the evaluation take place (tick all relevant; applicable response, n=25)?
Base/Proficiency checks…. 24% Line/Route checks…. 36%
Simulator/Line Oriented Flight Training….36% Other…. 4%

Number of times evaluation is carried on the occasions described above (Applicable responses, n= 9)
1…. 0% 2…. 22% 3…. 78%

19. Has the information on the CRM *skills* of flight crew been used to evaluate CRM training (applicable responses, n=9)?
Yes…. 56% No…. 44%

Comments specific to this section?
Yes…. 9% No…. 91%

Organisation

20. Have any organisational performance evaluations been carried out within the last two years (If no/don't know, go to question 23)?
Yes…. 36.4% No…. 36.6% Don't know…. 27%

21. What methods of evaluation were used (Tick all relevant; applicable responses, n=10)?

Company climate survey.....10% Business performance..... 0%

Safety performance ..... 30% Confidential reporting 20%

Incident reporting.....10% Technical performance 30%

Other..... 0%

Number of different evaluations used (applicable responses, n=12)
1....42% 2.... 16.5% 3....25% 4....0% 5....0% 6....16.5%

22. Has the information collected on the organisation been used to evaluate CRM training (applicable responses, n=12)?
Yes....33.3% No....58.3% Don't know....8.3%

Comments specific to this section?
Yes…. 9% No…. 91%

Other evaluation methods

23. Does your company use any other method to evaluate the effectiveness of CRM training, not described above? (If no/don't know, go to question 24)
Yes…. 0% No…. 100%

Reasons why CRM training effectiveness is not evaluated

24. What factors do you think prevent companies from evaluating CRM training effectiveness? (Tick all relevant; applicable responses; n= 32)

Time….22% Resources…. 22% Management support…. 15.5% Expertise…. 12.5%
Availability of measurement systems…. 22% Other…. 6%

Number of different factors chosen (Applicable responses, n= 11)

0…. 0% 1…. 18% 2…. 18% 3…. 27.5% 4…. 27.5% 5…. 9%

25. Do you think being able to evaluate CRM training effectiveness would be beneficial and if so why?

Yes…. 82% No…. 9% Don't know…. 0% No response…. 9%

Comments specific to this section?

Yes…. 45% No…. 55%

Methods to support CRM training evaluation effectiveness

26. What do you think are the preferred methods for evaluating whether CRM training has transferred to the flight deck? (Put the top **three** in order, with 1= best method)

Base/proficiency checks	**12**	Line/route checks	**13**
Simulator/LOFT checks	**20**	Incident reports	2
Accident data	0	Confidential report	2
Technical performance	4	Attitude surveys	7
Interviews sessions	0	Self/peer/360° assessment	0
Feedback questionnaire	1	Knowledge assessment	5
Other	0	No response	0

27. What in your opinion, are the main problems relating to the evaluation of CRM training effectiveness (Put the top **three** in order with 1= greatest problem)?

Availability of measurement systems	**19**	Quality of measurement systems	**9**
Competence/expertise of evaluators	**11**	Flight crews' attitudes/acceptance	**9**
Management	5	Financial	4
Other (Time)	3	No response	6

28. What could be done to help aviation companies to evaluate the effectiveness of their CRM programme?

Yes…. 64% No…. 36%

General comments

Yes…. 9% No…. 91%

Appendix 6 Questionnaire responses: medium carriers

Background

Response
AOC holders… 23 (68% response)

Type of licence
Type A…. 100%

Aircraft
Fixed-wing…. 91% Rotary…. 9%

Flight Crew CRM Training

1. Do flight crew employed by your organisation receive basic/foundation CRM training?
Yes…. 100% No…. 0%

2. Do flight crew employed by your organisation receive recurrent CRM training?
Yes…. 96% No…. 4%

3. Who provides the basic/foundation training?
In-house training department…. 74% Another airline…. 4.5%
Specialist consultancy…. 17% Other…. 4.5%

4. Who provides the recurrent training?
In-house training department…. 78% Another airline…. 0% Specialist consultancy…. 13%
Other …. 4.5% Not relevant…. 4.5%

5. If the training is not provided in-house, please give the name and contact details of the course provider.
Information held by researchers

6. Is the CRM training course designed for:
The specific operations carried out by your flight crew?…. 57% Generic ? …. 26%
Both…. 17%

Please add any further comments relevant to CRM training?
Yes…. 48% No…. 52%

Reactions

7. Is an assessment made of the *reactions* of flight crew to the CRM training?
Yes…. 57% No…. 39% Don't know…. 4%

8. How is this done? (Tick all relevant; applicable responses, n=14)
Reaction sheet…. 29% Oral feedback/debriefing…. 71%

Comments specific to this section?
Yes…. 43% No…. 57%

Attitudes

9. Has your organisation carried out an assessment of the *attitudes* of flight crew to the concepts covered in CRM training within the past two years?
Yes…. 43% No…. 57%

10. How was this done? (Tick all relevant; applicable responses, n=10)
Company specific attitude questionnaire….10%
Cockpit/ Flight Management Attitude Questionnaire (*CMAQ/FMAQ*)…. 10%
Other (informal oral feedback)…. 80%

11. Has this information on the *attitudes* of flight crew been used to evaluate CRM training (applicable responses, n=10)?
Yes…. 70% No…. 30%

Comments specific to this section?
Yes…. 30% No…. 70%

Knowledge

12. Has your airline carried out an assessment of the extent of flight crews' *knowledge* of the concepts covered in CRM training in the past two years? (If no/don't know, go to question 15)
Yes…. 26% No…. 70% No response…. 4%

13. How was this done (Tick all relevant; applicable responses, n=6)?
Multiple choice test…. 16.5% Written exam…. 16.5% Oral feedback…. 67%

14. Has the information on the CRM *knowledge* of flight crew been used to evaluate CRM training (applicable responses, n=6)?
Yes…. 100% No…. 0%

Comments specific to this section?
Yes…. 30% No…. 70%

Behaviour

15. Has your airline carried out an assessment of flight crews' CRM *skills* within the past two years? (If no/don't know, go to question 20)
Yes…. 70% No…. 30%

16. How was this done (% of applicable responses, n= 16)?
Behavioural marker system ….18.5% Technical checklist…. 13%
Informal feedback …. 68.5%

17. If you are using behavioural marker system, what do you use (Tick all relevant; applicable response, n=4)?
Company specific ….25% *NOTECHS* …. 75%

18. When does the evaluation take place (tick all relevant; applicable response, n=37)?
Base/Proficiency checks…. 32.3% Line/Route checks…. 32.3%
Simulator/Line Oriented Flight Training….32.3% Other…. 3%

Number of different occasions evaluation is carried out (applicable response, n=16)
1…. 31% 2…. 19% 3…. 50%

19. Has the information on the CRM *skills* of flight crew been used to evaluate CRM training (applicable responses, n=16)?
Yes…. 59% No…. 35% No response…. 6%

Comments specific to this section?
Yes…. 30% No…. 70%

Organisation

20. Have any organisational performance evaluations been carried out within the last two years? (If no/don't know, go to question 23)
Yes…. 52% No…. 39% Don't know …. 4.5% No response…. 4.5%

21. What methods of evaluation were used (Tick all relevant; applicable responses, n=30)?

Company climate survey.....10% Business performance.....17%

Safety performance ..... 23% Confidential reporting 13%

Incident reporting..... 23% Technical performance 7%

Other (all training audits)..... 7%

Number of different evaluations used (applicable responses, n=12)
1…. 42% 2…. 16.5% 3…. 25% 4…. 0% 5…. 0% 6…. 16.5%

22. Has the information collected on the organisation been used to evaluate CRM training (applicable responses, n=12)?
Yes…. 33.3% No…. 58.3% Don't know …. 8.3%

Comments specific to this section?
Yes…. 9% No…. 91%

Other evaluation methods

23. Does your company use any other method to evaluate the effectiveness of CRM training, not described above? (If no/don't know, go to question 24)

Yes…. 0% No…. 100%

Reasons why CRM training effectiveness is not evaluated

24. What factors do you think prevent companies from evaluating CRM training effectiveness? (Tick all relevant; applicable responses = 51)

Time...21.5% Resources... 23.5% Management support... 15.5% Expertise …13.5%
Availability of measurement systems…. 22% Other…. 4%

Number of different factors chosen (applicable responses = 23)

0…. 17% 1…. 13% 2…. 35% 3…. 9% 4…. 17% 5…. 9%

25. Do you think being able to evaluate CRM training effectiveness would be beneficial and if so why?

Yes…. 70% No…. 4% Don't know…. 4% No response…. 22%

Comments specific to this section?

Yes…. 57% No…. 43%

Methods to support CRM training evaluation effectiveness

26. What do you think are the preferred methods for evaluating whether CRM training has transferred to the flight deck? (Put the top **three** in order, with 1= best method)

Base/proficiency checks	8	Line/route checks	**38**
Simulator/LOFT checks	**36**	Incident reports	5
Accident data	1	Confidential report	8
Technical performance	0	Attitude surveys	6
Interviews sessions	4	Self/peer/360° assessment	6
Feedback questionnaire	9	Knowledge assessment	3
Other	1	No response	13

27. What in your opinion, are the main problems relating to the evaluation of CRM training effectiveness (Put the top **three** in order with 1= greatest problem)?

Availability of measurement systems	**34**	Quality of measurement systems	**27**
Competence/expertise of evaluators	**28**	Flight crews' attitudes/acceptance	16
Management	5	Financial	9
Other	0	No response	19

28. What could be done to help aviation companies to evaluate the effectiveness of their CRM programme?

Comments
Yes…. 52% No…. 48%

General Comments

Yes…. 22% No…. 78%

Appendix 7 Questionnaire responses: small carriers

Background

Response
AOC holders… 56 (57% response) PAOC holders…22 (73% response)

Type of licence
Type A…. 0% Type B…. 71% Police …. 29%

Aircraft
Fixed-wing…. 44% Rotary…. 56%

Flight Crew CRM Training

1. Do flight crew employed by your organisation receive basic/foundation CRM training?
Yes…. 96% No…. 3% No response …. 1%

2. Do flight crew employed by your organisation receive recurrent CRM training?
Yes…. 81% No…. 19%

3. Who provides the basic/foundation training?
In-house training department…. 23% Another airline…. 21% Specialist consultancy…50%
Other…. 2.5% No training provided…. 2.5% No response…. 1%

4. Who provides the recurrent training?
In-house training department…. 56.5% Another airline…. 9%
Specialist consultancy…. 15.5% Other …. 0% No training provided…. 19%

5. If the training is not provided in-house, please give the name and contact details of the course provider.
Information held by researchers

6. Is the CRM training course designed for:
The specific operations carried out by your flight crew?…. 40% Generic ? …. 49%
Both…. 9% Not applicable…. 1% No response…. 1%

Please add any further comments relevant to CRM training?
Yes…. 35% No…. 65%

Reactions

7. Is an assessment made of the *reactions* of flight crew to the CRM training?
Yes…. 56% No…. 39% Don't know….. 5%

8. How is this done? (Tick all relevant; applicable responses, n=43)
Reaction sheet…. 14% Oral feedback/debriefing…. 84% Other…. 0% No response…. 2%

Comments specific to this section?
Yes…. 15.5% No…. 84.5%

Attitudes

9. Has your organisation carried out an assessment of the *attitudes* of flight crew to the concepts covered in CRM training within the past two years?
Yes…. 15.4% No…. 83.6% No response…. 1%

10. How was this done? (Tick all relevant; applicable responses, n=12)
Company specific attitude questionnaire….8%
Cockpit/ Flight Management Attitude Questionnaire (*CMAQ/FMAQ*)…. 0%
Other (informal oral feedback)…. 92%

11. Has this information on the *attitudes* of flight crew been used to evaluate CRM training (applicable responses, n=12)?
Yes…. 75% No…. 25%

Comments specific to this section?
Yes…. 8% No…. 92%

Knowledge

12. Has your airline carried out an assessment of the extent of flight crews' *knowledge* of the concepts covered in CRM training in the past two years? (If no/don't know, go to question 15)
Yes…. 42% No…. 58%

13. How was this done (Tick all relevant; applicable responses- first method, n= 33; second method, n= 4)?

1st Method
Multiple choice test…. 12% Written exam…. 12% Oral feedback…. 76%

2nd Method
Oral feedback…. 100%

14. Has the information on the CRM *knowledge* of flight crew been used to evaluate CRM training (Applicable responses, n=33)?
Yes…. 18% No…. 82%

Comments specific to this section?
Yes…. 8% No…. 92%

Behaviour

15. Has your airline carried out an assessment of flight crews' CRM skills within the past two years? (If no/don't know, go to question 20)
Yes…. 44% No…. 55% Don't know…. 1%

16. How was this done (Applicable responses, n= 34)?
Behavioural marker system ….11.5% Technical checklist…. 24%
Informal feedback …. 64.5%

17. If you are using behavioural marker system, what do you use (Tick all relevant; applicable response, n= 4)?
Company specific ….50% Line/LOS checklist…. 50% NOTECHS …. 0%

18. When does the evaluation take place (tick all relevant; % of applicable response, n=56)?
Base/Proficiency checks…. 44% Line/Route checks…. 46%
Simulator/Line Oriented Flight Training….5% Other…. 5%

Number of times evaluation is carried on the occasions described above (Applicable responses, n= 34)
1…. 50% 2…. 32% 3…. 18%

19. Has the information on the CRM skills of flight crew been used to evaluate CRM training (applicable responses, n=34)?
Yes…. 41% No…. 56% No response…. 3%

Comments specific to this section?
Yes…. 10% No…. 90%

Organisation

20. Have any organisational performance evaluations been carried out within the last two years? (If no/don't know, go to question 23)

Yes…. 27% No…. 64% Don't know…. 8% No response…. 1%

21. What methods of evaluation were used (Tick all relevant; applicable responses, n=43)?

Company climate survey .....9% Business performance.....14%

Safety performance.....19% Confidential reporting 18%

Incident reporting .....26% Technical performance 5%

Other (all training audits) 9%

Number of different evaluations used (applicable responses, n=21)
1…. 38% 2…. 33% 3…. 14% 4…. 10% 5…. 5%

22. Has the information collected on the organisation been used to evaluate CRM training (applicable responses, n=21)?
Yes…. 47.5% No…. 47.5% Don't know…. 5%

Comments specific to this section?
Yes…. 4% No…. 96%

Other evaluation methods

23. Does your company use any other method to evaluate the effectiveness of CRM training, not described above? (If no/don't know, go to question 24)

Yes…. 1% No…. 99%

Please describe the method (Applicable responses, n=1)

360° appraisal…. 1%

Reasons why CRM training effectiveness is not evaluated

24. What factors do you think prevent companies from evaluating CRM training effectiveness? (Tick all relevant; applicable response = 150)

Time….28% Resources…. 26% Management support…. 7.5% Expertise…. 20%
Availability of measurement systems…. 17.5% Other…. 1%

Number of different factors chosen (Applicable response = 78)

0…. 15% 1…. 26% 2…. 26% 3…. 18% 4…. 12% 5…. 3%

25. Do you think being able to evaluate CRM training effectiveness would be beneficial and if so why?

Yes…. 46% No…. 18% Don't know…. 17% No response…. 19%

Comments specific to this section?

Yes…. 58% No…. 42%

Methods to support CRM training evaluation effectiveness

26. What do you think are the preferred methods for evaluating whether CRM training has transferred to the flight deck? (Put the top **three** in order, with 1= best method)

Base/proficiency checks	109	Line/route checks	129
Simulator/LOFT checks	33	Incident reports	15
Accident data	8	Confidential report	31
Technical performance	1	Attitude surveys	12
Interviews sessions	23	Self/peer/360° assessment	38
Feedback questionnaire	17	Knowledge assessmen	15
Other	3	No response	31

27. What in your opinion, are the main problems relating to the evaluation of CRM training effectiveness (Put the top **three** in order with 1= greatest problem)?

Availability of measurement systems	**127**	Quality of measurement systems	**72**
Competence/expertise of evaluators	**71**	Flight crews' attitudes/acceptance	43
Management	26	Financial.....	58
Other.....	3	No response	67

28. What could be done to help aviation companies to evaluate the effectiveness of their CRM programme?
Yes…. 42% No…. 54% No response…. 4%

General comments

Yes…. 27% No…. 73%

Appendix 8 The NOTECHS Behavioural Markers

Categories	Elements	Example Behaviours
Co-operation	Team building and maintaining	- Establishes atmosphere for open communication and participation
	Considering others	- Takes condition of other crew members into account
	Supporting others	- Helps other crew members in demanding situation
	Conflict solving	- Concentrates on what is right rather than who is right
Leadership & Managerial Skills	Use of authority & assertiveness	- Takes initiative to ensure involvement and task completion
	Maintaining standards	- Intervenes if task completion deviates from standards
	Planning and co-ordinating	- Clearly states intentions and goals
	Workload management	- Allocates enough time to complete tasks
Situation Awareness	System awareness	- Monitors and reports changes in systems states
	Environmental awareness	- Collects information about the environment
	Anticipation	- Identifies possible/ future problems
Decision Making	Problem definition / diagnosis	- Reviews causal factors with other crew members
	Option generation	- States alternative courses of action - Asks other crew member for options
	Risk assessment / Option choice	- Considers and shares risks of alternative courses of action
	Outcome review	- Checks outcome against plan

Rating Scale

Very Poor	Poor	Acceptable	Good	Very Good
Observed behaviour directly endangers flight safety	Observed behaviour in other conditions could endanger flight safety	Observed behaviour does not endanger flight safety but needs improvement	Observed behaviour enhances flight safety	Observed behaviour optimally enhances flight safety and could serve as an example for other pilots

Appendix 9 University of Texas Behavioural Markers

(Reprinted with permission of Professor Robert Helmreich)

The markers listed below are used in Line Operations Safety Audits, non-jeopardy observations of crews conducting normal line flights. Each of these markers has been validated as relating to either threat and error avoidance or management. With the exception of two global ratings, specific markers are rated (if observed) during particular phases of flight. Following is a list of currently used markers showing phase where rated, followed by the ratings for each phase of flight:

Key to Phase: P=Pre-departure/Taxi; T=Takeoff/Climb; D=Descent/Approach/Land; G=Global

			Phase
SOP Briefing	The required briefing was interactive and operationally thorough	- Concise, not rushed, and met SOP requirements - Bottom lines were established	P-D
Plans Stated	Operational plans and decisions were communicated and acknowledged	- Shared understanding about plans - "Everybody on the same page"	P-D
Workload Assignment	Roles and responsibilities were defined for normal and non-normal situations	- Workload assignments were communicated and acknowledged	P-D
Contingency Management	Crew members developed effective strategies to manage threats to safety	- Threats and their consequences were anticipated - Used all available resources to manage threats	P-D
Monitor / Cross-check	Crew members actively monitored and cross-checked systems and other crew members	- Aircraft position, settings, and crew actions were verified	P-T-D
Workload Management	Operational tasks were prioritized and properly managed to handle primary flight duties	- Avoided task fixation - Did not allow work overload	P-T-D
Vigilance	Crew members remained alert of the environment and position of the aircraft	- Crew members maintained situational awareness	P-T-D
Automation Management	Automation was properly managed to balance situational and/or workload requirements	- Automation setup was briefed to other members - Effective recovery techniques from automation anomalies	P-T-D
Evaluation Of Plans	Existing plans were reviewed and modified when necessary	- Crew decisions and actions were openly analyzed to make sure the existing plan was the best plan	P-T
Inquiry	Crew members asked questions to investigate and/or clarify current plans of action	- Crew members not afraid to express a lack of knowledge - "Nothing taken for granted" attitude	P-T
Assertiveness	Crew members stated critical information and/or solutions with appropriate persistence	- Crew members spoke up without hesitation	P-T
Communication Environment	Environment for open communication was established and maintained	- Good cross talk – flow of information was fluid, clear, and direct	G
Leadership	Captain showed leadership and coordinated flight deck activities	- In command, decisive, and encouraged crew participation	G

Rating Scale

1	2	3	4
Poor Observed performance had safety implications	**Marginal** Observed performance was barely adequate	**Good** Observed performance was effective	**Outstanding** Observed performance was truly noteworthy

References

Abelson, R. (1978) Are attitudes necessary? In B.T. King & E. McGinnies (Eds.) *Attitudes, conflict, and social change.* New York: Academic Press.

Alkov, R. A. (1989). The US Naval aircrew coordination program. In R. Jensen (Ed.) *Proceedings of the 5th International Symposium on Aviation Psychology,* Columbus, Ohio, 483-488.

Alkov, R. A. (1991). US Navy aircrew coordination training - a progress report. In R. Jensen (Ed.) *Proceedings of the 6th International Symposium on Aviation Psychology,* Columbus, Ohio, 368-371.

Alkov, R. A. & Gaynor, J. A. (1991). Attitude changes in Navy/Marine flight instructors following aircrew coordination training course. *International Journal of Aviation Psychology, 1(3),* 245-253.

Alliger, G. M. & Katzman, S. (1997). When training affects variability: beyond the assessment of mean differences in training evaluation. In J. K. Ford (Ed.), *Improving training effectiveness in work organizations* (pp. 223-246). Mahwah, NJ: Erlbaum.

Avermaete, v., J.A.G. & Kruijsen, E. (Eds.). (1998). *The evaluation of non-technical skills of multi-pilot aircrew in relation to the JAR-FCL requirements.* (Project report: CR-98443). Amsterdam, The Netherlands: NLR: EC NOTECHS.

Baker, D., Bauman, M. & Zalesny, M. D. (1991). Development of aircrew coordination exercises to facilitate training transfer. In R. Jensen (Ed.) *Proceedings of the 6th International Symposium on Aviation Psychology,* Columbus, Ohio.

Baker, D., Clothier, C., Woody, J. R., McKinney, E. H. & Brown, J. L. (1996). Crew Resource Management: A simulator study comparing fixed versus formed aircrews. *Aviation, Space, and Environmental Medicine, 67*(1), 3-7.

Baker, D. & Mulqueen, C. (1999). Pilot instructor/evaluator rater training: Guidelines for development. In R. Jensen (Ed.) *Proceedings of the 9th International Symposium on Aviation Psychology,* Columbus, Ohio, 332-337.

Baker, D., Mulqueen, C. & Dismukes, R.K. (2001). Training rater to assess resource management skills. In E. Salas & C. A. Bowers & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 131-146). Mahwah, NJ: Lawrence Erlbaum Associates.

Baker, D., Prince, C., Shrestha, L., Oser, R. & Salas, E. (1993). Aviation computer games for crew resource management training. *International Journal of Aviation Psychology, 3(2)*, 143-156.

Bannister, J. (2001). *Airline insurance: An introductory guide*.

Available: www.generali.com/library/airline/impagino.pdf [2001, 3/12/01].

British Airways. (2001). *British Airways Safety Information System (BASIS)*.

Available www.winbasis.com/basis/basis_about_1.htm [2001, 15/11/01].

Boehm-Davis, D. A., Holt, R. W. & Seamster, T. L. (2001). Airline resource management programs. In E. Salas & C. Bowers & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 191-216). Mahwah, NJ: Lawrence Erlbaum Associates.

British Psychological Society. (2001). *360 degree feedback: Best practice guidelines*. Leicester: British Psychological Society. Available www.bps.org.uk/documents/360bestprgdlns.pdf [2001, 14/12/01).

Brun, W., Eid, J., Johnsen, B.H., Ekornås, B., Laberg, J.C. & Kobbeltvedt, T. (2000) Shared mental models and task performance: Studying the effects of a crew and bridge resource management training program. (Project report: 1 2001). Amsterdam, Norway: Militaer Psykologi og Ledelse.

Butler, R. E. (1993). LOFT: Full mission simulation as CRM training. In E. Wiener & B. Kanki & R. Helmreich (Eds.), *Cockpit resource management* (pp. 230-258). San Diego: Academic Press.

Byrdorf, P. (1998). Human Factors and Crew Resource Management: An example of successfully applying the experience from CRM programmes in the Aviation World to the Maritime World. Paper presented at the 23rd Conference of the European Association for Aviation Psychology, Vienna.

Byrnes, R. E. & Black, R. (1993). Developing and implementing CRM programmes: The Delta experience. In E. Wiener, B. Kanki & R. Helmreich (Eds.), *Cockpit resource management* (pp. 421-443). San Diego: Academic Press.

CAA. (1995). *United Kingdom aeronautical information circular: crew resource management training*, AIC 37/1995. London: Civil Aviation Authority.

CAA. (1998a). *United Kingdom aeronautical information circular: crew resource management,* AIC 117/1998. London: Civil Aviation Authority.

CAA. (1998b). *United Kingdom aeronautical information circular: flight crew CRM training standards,* AIC 114/1998. London: Civil Aviation Authority.

Civil Aviation Authority. (2001). *Size of UK airlines by available capacity: April 2000-March 2001*. Available: www.caaerg.co.uk/ADUPublishedTables/Airline%20Data/200103/012_Size_of_UK_Airlines_in_seat_km_available_and_used.XLS.

Cannon-Bowers, J. A. & Salas, E. (1997). Teamwork competencies: The interaction of team member knowledge, skills and attitudes. In H. F. O'Neil (Ed.), *Workforce readiness: competencies and assessment.* Hillsdale, NJ: LEA.

Chidester, T. R., Helmreich, R., Gregorich, S. E. & Geis, C. E. (1991). Pilot personality and crew coordination: Implications for training and selection. *International Journal of Aviation Psychology, 1(1)*, 25-44.

Clark, R. E., Nielsen, R. A. & Wood, R. L. (1991). The interactive effects of cockpit resource management, domestic stress, and information processing in commercial aviation. In R. Jensen (Ed.) *Proceedings of the 6th International Symposium on Aviation Psychology,* Columbus, Ohio, 776-781.

Clothier, C. (1991). Behavioral interactions across various aircraft types: Results of systematic observations of line operations and simulations. In R. Jensen (Ed.) *Proceedings of the 6th International Symposium on Aviation Psychology*, Columbus, Ohio, 332-337.

Connelly, P. (1997). *A resource package for CRM developers: behavioural markers of CRM skills from real world case studies- and accidents. (Tech. Rep. No. 97-3)*. Austin: University of Texas, Aerospace Research Project.

Diehl, A. (1991a). Does cockpit management training reduce aircrew error? Paper presented at the 22nd International Seminar International Society of Air Safety Investigators, Canada.

Diehl, A. (1991b). The effectiveness of training programs for preventing aircrew 'error'. In R. Jensen (Ed.) *Proceedings of the 6th International Symposium on Aviation Psychology,* Columbus, Ohio, 640-655.

Dwyer, D. J., Fowlkes, J. E., Oser, R. L., Salas, E. & Lane, N. E. (1997). Team performance measurement in distributed environments: The TARGETs methodology. In M. T. Brannick, E. Salas & C. Prince (Eds.), *Team performance assessment and measurement. theory, methods and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Elliott-Mabey, N. (1999). *The assessment of RAF attitudes to CRM issues: a three year comparison* (PTC/496192/7/CSSB): RAF Command Scientific Branch.

FAA (1993). *Advisory circular 120-51A crew resource management training* Washington: US Department of Transport. Washington: FAA.

FAA (1998). *Advisory circular 120-51C: crew resource management training* Washington: US Department of Transport. Washington: FAA.

Flin, R. & Martin, L. (1998). *Behavioural markers for crew resource management. (Civil Aviation Authority*, Paper 98005. London: Civil Aviation Authority.

Flin, R. & Martin, L. (2001). Behavioural Markers for CRM: A review of current practice. *International Journal of Aviation Psychology, 11*, 95-118.

Flin, R. & O'Connor, P. (2001). Applying crew resource management on offshore oil platforms. In E. Salas, C. A. Bowers & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 217-234). Mahwah, NJ: Lawrence Erlbaum Associates.

Fonne, V. M. & Fredriksen, O. K. (1995). Resource management and crew training for HSV-navigators. In R. Jensen (Ed.) *Proceedings of the 8th International Symposium of Aviation Psychology*, Columbus, Ohio, 585-590.

Geis, C. E. (1987). Changing attitudes through training: a formal evaluation of training effectiveness. In R. Jensen (Ed.) *Proceedings of the 3rd International Symposium on Aviation Psychology,* Columbus, Ohio, 392-398.

George Mason University. (1996). *Improving Crew Assessments. Training materials to accompany a FAA sponsored workshop on evaluator calibration*. Washington: George Mason University.

Goeters, K. M. (2000). Validation of CRM training by NOTECHS: Results from the PHARE ASI project. Paper presented at the 24th Conference of the European Association for Aviation Psychology, Crieff, Scotland.

Gregorich, S. E., Helmreich, R. L. & Wilhelm, J. A. (1990). The structure of cockpit management attitudes. *Journal of Applied Psychology, 75 (6)*, 682-690.

Gregorich, S. E. & Wilhelm, J. A. (1993). Crew resource management training assessment. In E. L. Wiener & B. G. Kanki & R. L. Helmreich (Eds.), *Cockpit Resource Management* (pp. 173-196): San Diego: Academic Press.

Gregorich, S. E. (1993). The dynamics of CRM attitude change: Attitude stability. In R. Jensen (Ed.) Proceedings of the 7th International Symposium of Aviation Psychology, Columbus, Ohio, 509-512.

Grubb, G., Morey, J. C. & Simon, R. (2001). *Sustaining and advancing performance improvements achieved by crew resource management training.* Paper presented at the 11th International Symposium of Aviation Psychology, Columbus, Ohio, April.

Hackman, J. R. (1983). *A normative model of work team effectiveness*. Hew Haven, CT: Yale University.

Hamblin, A. C. (1974). *Evaluation and control of training*. London: McGraw Hill.

Hansberger, J. T., Holt, R. W. & Boehm-Davies, D. (1999). Instructor/evaluator evaluations of ACRM effectiveness. In R. Jensen (Ed.) *Proceedings of the 10th International Symposium on Aviation Psychology*, Ohio State University, Columbus, OH.

Harrington, D. K. & Kello, J. E. (1992, June 1992). Systematic evaluation of nuclear operator team skills training: A progress report. *Paper presented at the STL conference on Human Factors and Power plants*, Monterey, California.

Harris, J. S. (1995). Crew resource management applies to single-pilot flight operations. *Flight Safety Foundation: Helicopter safety, 21*(5), 1-4.

Häusler, K. (2000) CRM behaviour and team performance under high workload: Outline and implications of a simulator study. *In Proceedings of the Australian Aviation Psychology Symposium*, Manley, Australia.

Haynes, A. (1992) United 232: Coping with the 'one-chance-in-a-billion'' loss of all flight controls. *Flight Deck, 3*, Spring, 5-21.

Hayward, B. & Alston, N. (1991). Team building following a pilot labour dispute: extending the CRM envelope. In R. Jensen (Ed.) *Proceedings of the 6th International Symposium on Aviation Psychology*, Columbus, Ohio, 377-383.

Health and Safety Executive. (1999). *Reducing error and influencing behaviour. HSG48*. Suffolk: HSE Books.

Heiman, G. A. (1995). *Research methods in psychology*. Boston, MA: Houghton Mifflin Co.

Helmreich, R. (1984). Cockpit management attitudes. *Human Factors, 26*, 583-589.

Helmreich, R. L. & Foushee, H. C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E. L. Wiener & B. G. Kanki & R. L. Helmreich (Eds.), *Cockpit Resource Management* (pp. 3-41). New York: Academic Press.

Helmreich, R. L., Klinect, J. R. & Wilhelm, J. A. (1999*). The line operations safety audit (LOSA) observer's manual, version 7.0*. (Tech. Rep. 99-0). Austin, TX: NASA/ University of Texas/Federal Aviation Administration Aerospace Group.

Helmreich, R. L., Merritt, A. C. & Wilhelm, J. A. (1999). The evolution of Crew Resource Management training in commercial aviation. *International Journal of Aviation Psychology, 9*, 19-32.

Helmreich, R., Wilhelm, J., Kello, J., Taggart, E. & Butler, R. (1990). *Reinforcing and evaluating crew resource management: Evaluator/LOS instructor manual*. Austin: NASA/UT/FAA Aerospace Group.

Helmreich, R. (2000). *The Line Operations Safety Audit (LOSA). (Version 9)*. Austin: NASA/University of Texas/Federal Aviation Administration Aerospace Group.

Helmreich, R. L. (1996). *The evolution of Crew Resource Management.* Paper presented at the IATA Human Factors Seminar, Warsaw, Poland.

Helmreich, R. L., Klinect, J. R. & Wilhelm, J. A. (2000). Managing threat and error: Data from line operations. *In Proceedings of the Australian Aviation Psychology Symposium*, Manley, Australia.

Helmreich, R. L. & Wilhelm, J. A. (1991). Outcomes of crew resource management training. *International Journal of Aviation Psychology, 1*, 287-300.

Henderson, S. (2000) Using assessment data- bridging the gap between checking and training. Paper presented at the Australian Aviation Psychology Conference, Manly, 20-24 November.

Hollnagel, E. (1993). *Human reliability analysis: context and control*. London: Academic Press.

Holt, R. W., Boehm-Davis, D. & Hansberger, J. T. (1999). Evaluating effectiveness of ACRM using LOE and line-check data. In R. Jensen (Ed.) *Proceedings of the 10th International Symposium on Aviation Psychology*, Ohio State University, Columbus, OH, 273-288.

Holt, R. W., Johnson, P. J. & Goldsmith, T. E. (2000). *Application of psychometrics to the calibration of air carrier evaluators*. Available: www.faa.gov/avr/afs/rhotlpap.pdf.

Holt, R. W., Boehm-Davis, D. A. & Beaubien, M. J. (2001). Evaluating resource management training. In E. Salas & C. Bowers & E. Edens (Eds.), *Improving teamwork in organizations: Applications of resource management training* (pp. 165-190). Mahwah, NJ: Lawrence Erlbaum Associates.

Howard, S. K., Gaba, D. M., Fish, K. J., Yang, G. S. & Sarnquist, F. H. (1992). Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. *Aviation, Space, and Environmental Medicine, 63*(9), 763-770.

Howell, D.C. (1992) *Statistical methods for psychology*. Belmont, CA: Wadsworth Inc.

Human Factors Group of the Royal Aeronautical Society. (1996). *Quality crew resource management*. London: RAeS.

Ikomi, P. A., Boehm-Davis, D., Holt, R. W. & Incalcaterra, K. A. (1999). Jump seat observations of Advanced Crew Resource Management (ACRM) effectiveness. In R. Jensen (Ed.) *Proceedings of the 10th International Symposium on Aviation Psychology*, Ohio State University, Columbus, OH, 292-297.

Incalcaterra, K. A. & Holt, R. W. (1999). Pilot evaluation of ACRM programs. In R. Jensen (Ed.) *Proceedings of the 10th International Symposium on Aviation Psychology*, Ohio State University, Columbus, OH, 285-291.

Irwin, C. M. (1991). The impact of initial and recurrent cockpit resource management training on attitudes. In R. Jensen (Ed.) *Proceedings of the 6th International Symposium on Aviation Psychology*, Columbus, Ohio, 344-349.

JAA. (2001). *JAR-OPS, 1 amendment 3 (November 2001), Crew resource management- flight crew*. Hoofddorp, Netherlands: EC Joint Aviation Authorities.

JAA (2001) *JAR-OPS 1, Change 3 (November 2001).Subpart N, section 1: Flight crew*. Hoofddorp, The Netherlands: Author.

Jentsch, F., Bowers, C. A. & Holmes, B. (1995). The acquisition and decay of aircrew coordination skills. In R. Jensen (Ed.) *Proceedings of the 8th International Symposium of Aviation Psychology*, Columbus, Ohio.

Kayten, P. J. (1993). The accident investigator's perspective. In E. Wiener, B. Kanki & R. Helmreich (Eds.), *Cockpit resource management* (pp. 199-230). San Diego: Academic Press.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig & L. R. Bittel (Eds.), *Training and development handbook*. New York: McGraw Hill.

Kirkpatrick, D. L. (1998). *Evaluating training programs*. San Fransisco: Berrett-Koehler.

Klampfer, B., Flin, R. Helmreich, R, Häusler, R., Sexton, B., Fletcher, G., Field, P., Staender, S., Lauche, K., Dieckmann, P. & Amacher, A. (2001) *Enhancing performance in high risk environments: Recommendations for the use of behavioural markers*. Report from the behavioural markers workshop, Zürich, June. Berlin: Damler Benz Foundation. Available on the internet from www.psyc.abdn.ac.uk/homedir/poconnor/GIHRE21.pdf

Kraiger, K., Ford, J. K. & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcome to new method of training evaluation. *Journal of Applied Psychology, 78(2)*, 311-328.

Lauber, J. K. (1984). Resource Management in the cockpit. *Air Line Pilot, 53*, 20-23.

Law, J.R. & Sherman, P.J. (1995). Do rater agree? Assessing inter-rater agreement in the evaluation of air crew resource management skills. In R. Jensen (Ed.) *Proceedings of the 8th International Symposium of Aviation Psychology*, Columbus, Ohio, 601-612.

Leedom, D. K. & Simon, R. (1995). Improving team coordination: a case for behavior -based training. *Military Psychology, 7*(2), 109-122.

Margerison, C., Davies, R. & McCann, D. (1987). High-flying management development. *Training and Development Journal, 41(2)*, 38-41.

Maschke, P., Goeters, K. M., Hormann, H. J. & Schiewe, A. (1995). The development of the DLR/Lufthansa crew resource management training program. In N. Johnson, R. Fuller & N. McDonald (Eds.) *Proceedings of the 21st conference of the European Association for Aviation Psychology*, 23-29. Aldershot: Avebury Aviation

Maurino, D., Reason, J., Johnston, N. & Lee, R. (1995). *Beyond aviation human factors*. Aldershot: Avebury.

McDougall, W. A., Seamster, T. L. & Edens, E. S. (1993*).* Instructors strategies in the assessment of aircrews: heuristics in CRM assessment and their operational implications*. Paper presented at the 37th Human Factors and Ergonomics Society Annual Meeting*, Santa Monica, CA.

Mulqueen, C. & Baker, D. P. (1999). Assessing I/E rater training effectiveness: issues in measurement. In R. Jensen (Ed.) *Proceedings of the 9th International Symposium on Aviation Psychology*, Columbus, Ohio, 323-328.

Naef, W. (1995). Practical applications of CRM concepts: Swissair's human aspects development program (HAD). In R. Jensen (Ed.) *Proceedings of the 8th International Symposium of Aviation Psychology*, Columbus, Ohio, 597-602.

O'Connor, P. & Flin, R. (under review). Crew resource management training for offshore production teams.

O'Connor, P., Hoermann, H. J., Flin, R., Goeters, K. M., Lodge, M. & the JARTEL group. (in press). Developing a method for evaluating CRM skills: A European perspective. *International Journal of Aviation Psychology*.

O'Leary, M. (1999). *The British Airways human factors reporting programme.* Paper presented at the Human Error, Safety, and System Development Conference, Liege, June.

Predmore, S. C. (1991). Microcoding of communication in accident investigation: crew coordination in United 811 and United 232. In R. Jensen (Ed.) *Proceedings of the 6th International Symposium on Aviation Psychology*, Columbus, Ohio, 350-355.

RAeS HFG (1999). Discussion document: Crew Resource Management, London: RAeS.

Royal Aeronautical Society, Human Factors Group. (1999). Discussion document: Crew Resource Management. London: RAeS.

Robertson, M. M. & Taylor, J. C. (1995). A systematic training evaluation model applied to measure the effectiveness of an aviation maintenance team training program. In R. Jensen (Ed.) *Proceedings of the 8th International Symposium of Aviation Psychology*, Columbus, Ohio, 631-636.

Salas, E. & Cannon-Bowers, J. (1995). Methods, tools and strategies for team training. In M. A. Quinones & A. Dutta (Eds.), *Training for 21st century technology: applications of psychological research*. Washington, D.C.: Jossey-Bass.

Salas, E. & Cannon-Bowers, J. A. (1997). Methods, tools and strategies for team training. In M. Quinones & E. Ehrestein (Eds.), *Training for a rapidly changing workplace: Applications in psychological research* (pp. 291-322). Washington DC: American Psychological Association Press.

Salas, E., Fowlkes, J. E., Stout, R. J., Milanovich, D. M. & Prince, C. (1999). Does CRM training improve teamwork skills in the cockpit? Two evaluation studies. *Human Factors, 41(2)*, 326-343.

Salas, E., Burke, C. S., Bowers, C. A. & Wilson, K. A. (in press). Team training in the skies: Does Crew Resource Management (CRM) training work? *Human Factors*.

Salas, E., Bowers, C. A. & Edens, E. (Eds.). (2001). *Improving teamworking in organizations: Applications of resource management training*. Mahwah: Lawrence Elbaum Associates.

Schein, E. H. (1988). *Organizational psychology*. Englewood Cliffs, NJ: Prentice-Hall Inc.

Schiewe, A. (1995). On the acceptance of CRM-methods by pilots: results of a cluster analysis. In R. Jensen (Ed.) *Proceedings of the 8th International Symposium of Aviation Psychology*, Columbus, Ohio, 540-545.

Seamster, T. L., Edens, E. S. & Holt, R. W. (1995). Scenario event sets and the reliability of CRM assessment. In R. Jensen (Ed.) *Proceedings of the 8th International Symposium of Aviation Psychology*, Columbus, Ohio, 613-618.

Sinclair, M. A. (1992). Subjective assessment. In J. R. Wilson & N. E. Corlett (Eds.), *Evaluation of human work: a practical ergonomics methodology* (pp. 58-88). London: Taylor & Francis.

Stout, R. J., Salas, E. & Kraiger, K. (1996). The role of trainee knowledge structures in aviation psychology. *International Journal of Aviation Psychology, 7*(3), 235-250.

Taggart, W. R. & Butler, R. E. (1989). CRM validation program. In R. Jensen (Ed.) *Proceedings of the 5th International Symposium on Aviation Psychology*, Columbus, Ohio, 468-475.

Taylor, J. C. (1998 May). *Evaluating the effectiveness of maintenance resource management (MRM)*. Paper presented at the 12th International Symposium on Human Factors in Aviation Maintenance, Canada: Vancouver.

Taylor, J. C. (1998). *Evaluating the effectiveness of Maintenance Resource Management (MRM)*. Paper presented at the 12th International Symposium on Human Factors in Aviation Maintenance, Washington, DC, FAA, August.

Taylor, J.C. (2000a) *A new model for measuring return on investment (ROI) for safety programs in aviation: An example from airline maintenance resource management (MRM).* In proceedings of the Advances in Aviation Safety Conference, Paper number 2000-01-2090, Daytona Beach, Florida.

Taylor, J. C. (2000b). Reliability and validity of the maintenance resources management/ technical operations questionnaire. *International Journal of Industrial Ergonomics, 26*(2), 217-230.

Vandermark, M. J. (1991). Should flight attendants be included in CRM training? A discussion of a major air carrier's approach to total crew training. *International Journal of Aviation Psychology, 1(1)*, 87-94.

Wagenaar, W. A. & Groenewg, J. (1987). Accidents at sea: multiple causes and impossible consequences. *International Journal of Man-Machine Studies, 27*, 587-598.

Wiener, E., Kanki, B. & Helmreich, R. (1993). *Cockpit Resource Management*. San Diego: Academic Press.

Williams, D. M., Holt, R. W. & Boehm-Davies, D. A. (1997). Training for inter-rater reliability: baselines and benchmarks. In R. Jensen (Ed.) *Proceedings of the 9th International Symposium on Aviation Psychology*, Columbus, Ohio, 541-549.

Woldring, M. & Isaac, A. (1999). *Team Resource Management test and evaluation*

(HUM.ET1.ST10.2000-REP-01) Brussels: EUROCONTROL.

Yamamori, H. & Mito, T. (1993). Keeping CRM is keeping the flight safe. In E. Wiener, B. Kanki & R. Helmreich (Eds.), *Cockpit resource management* (pp. 199-230). San Diego: Academic Press.